Title
Investigation of hemispheric asymmetry in reasoning with HD-tDCS and fMRI

Author
Marinsek, Nicole

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Santa Barbara

Investigation of hemispheric asymmetry in reasoning with HD-tDCS and fMRI

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Dynamical Neuroscience

by

Nicole L. Marinsek

Committee in charge:

Professor Michael Miller, Chair

Professor Michael Gazzaniga

Professor Scott Grafton

June 2018

The dissertation of Nicole Marinsek is approved.

Scott Grafton

Michael Gazzaniga

Michael Miller, Committee Chair

June 2018

Investigation of hemispheric asymmetry in reasoning with HD-tDCS and fMRI

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research advisor, Dr. Michael Miller. Not only has he always been encouraging, supportive, and helpful, but he also gave me the freedom to pursue my own research interests. I genuinely could not have asked for a better advisor and I am deeply grateful to be a part of his lab.

I would also like to thank my committee members: Dr. Scott Grafton taught me the importance of using patient studies to inform cognitive theories and Dr. Gazzaniga's research on split-brain patients provided both the inspiration and the theoretical foundation for these experiments.

I am grateful to all the members of the Miller lab, both past and present: Evan Layher, Tyler Santander, Justin Kantner, Danielle King, Amy Frithsen, Brian Lopez, Jeanne Li, and Misty Schubert. I'd especially like to thank Ben Turner who taught me how to analyze fMRI data when I first started graduate school and who has been a helpful resource and collaborator ever since.

I am thankful for the Miller Lab research assistants who helped collect data for these experiments, including Anjali Dixit, Saleem Omary, Thien Truong, Nohemi Reyes Meraz, Shelsea De La O, Omeed Soltanalipour, Will Mehring, and Isaiah Rodriguez-Anguiano. I would also like to extend a special thanks to Yanelliz Melchor Martinez who contributed to the creation of the stimuli used in the second experiment.

I am eternally grateful for the love and support from my family. My mom, brother, and grandma in particular are incredible blessings and strong influences in my life. Most of all, I would like to thank my husband, Stephen. He is my constant source of joy, comfort, and strength, and I would need to write a second dissertation to adequately describe how much he means to me and how grateful I am of him.

VITA OF NICOLE MARINSEK

marinsek@dyns.ucsb.edu
updated May 2018

EDUCATION

DOCTORAL CANDIDATE IN DYNAMICAL NEUROSCIENCE
University of California, Santa Barbara, 2012 – 2018

B.S. IN BIOLOGY & PSYCHOLOGY, MINOR IN CHEMISTRY
New Mexico Institute of Mining and Technology, 2008 – 2012

RESEARCH

GRADUATE STUDENT RESEARCHER
Dr. Michael Miller's Memory and Neuroimaging Lab
University of California, Santa Barbara, 2012 – Present
Use behavioral, neuroimaging, neurostimulation, and analytical techniques to study the neural dynamics of causal inference making and belief updating.

RESEARCH ASSISTANT
Dr. Snezna Rogelj's Drug Discovery Lab
New Mexico Institute of Mining and Technology, 2010 – 2012
Screened novel compounds for antifungal activity against Candida albicans and analyzed fungus morphology under various conditions.

RESEARCH INTERN
Dr. Floh Thiels Lab
Center for Neuroscience, University of Pittsburgh, Summer 2011
Helped compile behavioral and neurobiological profiles of rats that were exposed to stress in order to establish an animal model of depression for the lab.

TEACHING & MENTORSHIP

INSTRUCTOR OF RECORD FOR COGNITIVE PSYCHOLOGY
Psychological and Brain Sciences, UCSB, Summer 2016
Designed all aspects of the course, including the course's structure, content, policies, assignments, and exams. Gave lectures to a class of 67 undergraduates three times weekly and received glowing feedback.

TEACHING ASSISTANT
Psychological and Brain Sciences, UCSB, 2013 - 2014
Held office hours, tutored students, proctored and graded exams, graded homework assignments, and led lab sessions for Advanced Research Methods, Cognition (2 quarters), and Cognitive Neuroscience.

MENTOR

Academic Research Consortium, UCSB, 2017

Research Mentorship Program, UCSB, 2013 - 2014

Mentored a total of four high school students and one undergraduate student over three six-week periods. Helped them design a research study, collect and analyze data, and present their research findings in a paper, poster, and presentation.

WORK

DATA SCIENTIST

Evidation Health, Santa Barbara, 2018

Lead efforts to develop a new clinical outcome assessment to submit to the FDA by conducting background research, deploying a survey to over 2 million people, and analyzing their responses.

LEARNING COACH

Introduction to Chemistry, NMT, Fall 2011

Co-led biweekly lab sessions, tutored students, led a homework discussion group, and graded homework sets for students identified as being at-risk.

RESIDENT ASSISTANT

New Mexico Institute Of Mining And Technology, 2010 - 2012

BOARD CERTIFIED PHARMACY TECHNICIAN

Walgreen's Pharmacy, 2009 – 2013

AWARDS & FELLOWSHIPS

KAVLI SUMMER INSTITUTE IN COGNITIVE SCIENCE FELLOW

Santa Barbara, CA, 2016

ATTENDEE AND PRESENTER AT THE NOBEL LAUREATE CONFERENCE

65th annual Nobel Laureate Conference, Lindau, Germany, 2015

Selected by three committees to attend the conference and by Nobel Laureate Kurt Wuthrich to present at a Masterclass session. Awarded travel costs.

GRADUATE RESEARCH PROGRAM FELLOW

National Science Foundation, 2014 - Present

Awarded three years of support to pursue proposed research goals.

DOCTORAL SCHOLAR

Graduate Student Division, UCSB, 2012 - Present

Awarded two years of support.

GRADUATE STUDENT FELLOW

Sage Center For The Study Of The Mind, UCSB, 2012

Awarded a $5,000 scholarship.

MACEY SCHOLAR
New Mexico Institute Of Mining And Technology, 2012
Awarded a $5,000 scholarship.

PSYCHOLOGY DEPARTMENTAL AWARD
New Mexico Institute Of Mining And Technology, 2012

ANALYTICAL AWARD
American Chemistry Society, 2010

SKILLS

LANGUAGES
Python / MATLAB / Psychtoolbox / Bash / C / HTML / CSS

SOFTWARE
FSL / Caret / Freesurfer / Osirix / EEGlab / ERPlab

COURSES

RELVANT GRADUATE COURSEWORK
Memory / Statistics / Neuroanatomy / Patient Studies / Neuroimaging / Computational Neuroscience / Cerebral Cortex / Cognitive Neuroscience / fMRI Data Analysis / Linear Systems / Ordinary Differential Equations / EEG

ONLINE
Statistical Learning (Stanford) / Python (Codecademy) / Data Science (Udacity) / Machine Learning (Coursera, Stanford) / SQL (Stanford)

PAPERS

Marinsek, N., Gazzaniga, M. (2016). A Split-Brain Perspective on Illusionism. Journal of Consciousness Studies, 23(11-12), 149-159.

Turner, B. O., Marinsek, N., Ryhal, E., & Miller, M. B. (2015). Hemispheric lateralization in reasoning. Annals of the New York Academy of Sciences, 1359(1), 47-64.

Marinsek, N., Turner, B.O., Gazzaniga, M., & Miller, M.B. (2014). Divergent hemispheric reasoning strategies: reducing uncertainty versus resolving inconsistency. Frontiers in Human Neuroscience, 8(839), 1- 10. doi: 10.3389/fnhum.2014.00839

CHAPTERS

Colvin, M. M., Marinsek, N. L., Miller, M. B., & Gazzaniga, M. S. (2017). Split-brain Cases. The Blackwell Companion to Consciousness, 634-647.

Marinsek, N. L., Gazzaniga, M. S., & Miller, M. B. (2016). Split-brain, split-mind. In The Neurology of Consciousness (Second Edition) (pp. 271-279).

Colvin, M., Marinsek, N., Miller, M. B., & Gazzaniga, M. S (2015) Split-brain cases. Blackwell Companion to Consciousness. Chapter under review.

TALKS

Marinsek, N. (March 2017) Establishing causal relationships between hemispheric asymmetry and reasoning biases with tDCS. Dynamical Neuroscience departmental seminar, UCSB.

Marinsek, N. (November 2015) Tracking the neural dynamics of belief updating. Dynamical Neuroscience departmental seminar, UCSB.

Marinsek, N. (July 2015). Functional Magnetic Resonance Imaging. 65th annual Nobel Laureate Conference, Lindau, Germany.

Marinsek, N. (May 2014). Hypothesis formation and evaluation in the brain. 38th annual mini-conference for the Psychological and Brain Sciences Department, University of California, Santa Barbara, CA.

POSTERS

Marinsek, N. & Miller, M. B. (2018, March). tDCS-induced hemispheric asymmetry alters belief updating. Cognitive Neuroscience Society conference, Boston, MA.

Marinsek, N. & Miller, M. B. (2017, March). Using realistic jury scenarios to study belief updating. Poster presented at the annual Cognitive Neuroscience Society conference, San Francisco, CA.

Marinsek, N., Turner, B. O., Layher, E., & Miller, M. B. (2016, March). Hemispheric Asymmetry in Reasoning: A review and meta-analysis. Poster presented at the annual Cognitive Neuroscience Society conference, New York, NY.

Marinsek, N., Turner, B. O., Steindam, C., & Miller, M. B. (2015, March). Tracking the neural dynamics of hypothesis evaluation. Poster presented at the annual Cognitive Neuroscience Society conference, San Francisco, CA.

Marinsek, N., Turner, B. O., & Miller, M. B. (2014, August). A Bayesian approach to identifying the neural correlates of hypothesis formation and evaluation. Poster presented at the annual Bay Area Memory Meeting, Stanford University, Palo Alto, CA.

Marinsek, N., Turner, B. O., & Miller, M. B. (2014, April). Components of belief updating and their neural correlates. Poster presented at the annual Cognitive Neuroscience Society Meeting, Boston, MA.

Marinsek, N., Turner, B. O., & Miller, M. B. (2013, November). The neural correlates of hypothesis formation and evaluation. Poster presented at the annual Society for Neuroscience Meeting, San Diego, CA.

Marinsek, N., Turner, B. O., & Miller, M. B. (2013, August). The neural correlates of hypothesis formation and evaluation. Poster presented at the annual Bay Area Memory Meeting, San Francisco, CA.

ABSTRACT

Investigation of hemispheric asymmetry in reasoning with HD-tDCS and fMRI

by

Nicole Marinsek

Studies on multiple patient groups suggest that reasoning has a hemispheric asymmetry component. Previously, we proposed that neural networks in the left hemisphere are driven toward increasing and maintaining certainty, while right frontal networks prioritize congruence between beliefs and evidence. We tested the predictions of this framework with two high definition transcranial direct current stimulation (HD-tDCS) experiments and one functional magnetic resonance imaging (fMRI) experiment. In both HD-tDCS studies, we aimed to induce (or amplify) hemispheric asymmetry in healthy participants as they completed novel reasoning tasks. Each participant completed three tDCS sessions: a LH-bias session, in which the anode was placed over the left inferior frontal gyrus (IFG; BA45) and the cathode over the right IFG; a RH-bias session, in which the anode was placed over the right IFG and the cathode over the left IFG; and a sham session, which served as a control.

In the first HD-tDCS experiment, participants (N=26) completed a probabilistic inference task that required the integration of evidence and one's prior background knowledge. Consistent with predictions, we found that the intensity of RH-bias stimulation was associated with 1) collecting more evidence, 2) adopting a higher threshold for stopping evidence collection, and 3) making less certain guesses than an ideal Bayesian updater during the evidence presentation. Contrary to

predictions, we found that greater LH-bias intensity was associated with more evidence collection, and LH-bias stimulation was associated with greater belief backtracks after encountering conflicting evidence than RH-bias or sham stimulation.

The second HD-tDCS experiment followed a similar stimulation protocol but used reasoning problems that were more deeply embedded in real-world contexts in order to create more salient belief-evidence conflicts. During each stimulation session, 24 participants 1) judged whether a criminal suspect was guilty or not guilty based on crime scene evidence, 2) judged whether or not to pass a law based on arguments in favor and in opposition to it, and 3) judged whether a news headline was real or fake. We found that RH-bias stimulation reduced belief polarization after conflict, which was consistent with our predictions. Similarly, when evidence conflicted participants' strong beliefs, they backtracked on their beliefs more under RH-bias stimulation compared to sham stimulation and, albeit to a lesser extent, compared to LH-bias stimulation. Under RH-bias stimulation, participants were less likely to judge real news headlines as being real, which resulted in poorer discrimination of real vs. fake headlines compared to sham and LH-bias stimulation.

Finally, in the fMRI experiment, we examined lateralization in frontal anatomical regions for contrasts that we predicted to be more left-lateralized or more right-lateralized. Participants (N=36) completed a modified version of the state guessing task that was used in the first tDCS experiment. Consistent with predictions, contrasts involving uncertainty and belief advances were generally more left-lateralized and contrasts involving conflicting evidence and belief backtracks were more right-lateralized.

We show that HD-tDCS can alter belief updating in healthy individuals in a way that is consistent with the patient literature, but additional experiments are necessary to disentangle the causal relationships between different reasoning biases and neural activity in left and right frontal neural networks.

TABLE OF CONTENTS

I. Introduction

A rich history of patient studies reveal that different reasoning deficits arise when the left vs. the right frontal cortices are damaged or disconnected and suggest that human reasoning may have a hemispheric lateralization component. However, there is less convincing evidence of hemispheric asymmetry from studies on healthy individuals, principally because very few reasoning studies directly examine laterality. In this section, we review the evidence for hemispheric asymmetry in reasoning from studies on split-brain patients, patients with unilateral brain damage, and patients with delusional disorders, and the relatively weaker evidence from studies on healthy individuals. We synthesize these bodies of literatures into a simple framework and propose two neurostimulation experiments and one neuroimaging experiment that test the predictions of this framework in healthy individuals.

A. Evidence of hemispheric asymmetry in reasoning

1. Reasoning in the split-brain

Patients with severe epilepsy occasionally have their corpus callosum cut to isolate seizure activity to one hemisphere. This procedure reduces the frequency and severity of the seizures, but it leaves a patient with two divided, independent hemispheres – a split brain. Split-brain patients offer a unique opportunity to study the two hemispheres in isolation; by presenting stimuli to one half of the visual field, the mental capabilities of each hemisphere can be characterized independently of the other.

Using this method, Gazzaniga and colleagues discovered that the left hemisphere has a propensity to make inferences and create explanations. He named this tendency of the left hemisphere to create explanations the Left Brain Interpreter (Gazzaniga, 2000). In one experiment, a split-brain patient's left hemisphere was shown a picture of a chicken claw and his right hemisphere was shown a picture of a snow scene. The patient was asked to point to a card that was associated with the picture he just saw. With his left hand (controlled by his right hemisphere) he selected a shovel, which matched the snow scene, and with his right hand (controlled by his left hemisphere) he selected a chicken, which matched the chicken claw. The experimenter asked the patient why he selected each item. One would expect the speaking left hemisphere to explain why it chose the chicken but not why it chose the shovel, since the left hemisphere did not have access to that information. Instead, the patient's speaking left hemisphere replied, "Oh, that's simple. The chicken claw goes with the chicken and you need a shovel to clean out the chicken shed" (Gazzaniga, 2000). The left hemisphere quickly created an explanation for the behavior – an explanation that was incorrect but nonetheless plausible, given the left hemisphere's limited information. In another experiment, researchers presented the command to stand to the right hemisphere. After the patient stood, experimenters asked the patient why he did so. Again, instead of admitting that he did not know why he stood, the speaking left hemisphere created an explanation, insisting he was thirsty and wanted a drink (see Gazzaniga & Miller, 2009).

Although these anecdotes illustrate the interpretative abilities of the left hemisphere, they do not necessarily suggest that inference making is left lateralized; indeed, the right hemisphere may also readily create causal explanations, but these explanations go undetected because they are not verbalized. To directly test the inferential capabilities of each hemisphere, Gazzaniga and Smylie (1984) presented two pictures to one hemisphere of split-brain patients and asked the patients to point to a picture that depicted the outcome of combining the objects with their contralateral hand. For example, in one trial a patient's left hemisphere was shown pictures of a match and wooden log, and the patient could choose between pictures of a woodpile, a lit cigarette, or a bonfire (the correct causal outcome) with their right hand. This task required patients to infer the causal relationship between two objects. Even though both hemispheres possessed a sophisticated lexicon, only the left hemisphere could infer the causal relationship between the two items, and this trend held for both verbal and visual stimuli. Consistent with these results, Roser et al. (2005) showed that the left hemisphere can use evidence to extract an underlying causal structure, but the right hemisphere performs at chance. There's also evidence that the left hemisphere, but not the right hemisphere, may attempt to infer and predict an underlying causal structure even when told that events are random (Wolford, Miller, & Gazzaniga, 2000).

The left hemisphere's tendency to make inferences is also apparent in the realm of memory. In one experiment, Phelps and Gazzaniga (1992) presented visual scenes with a common theme to split-brain patients and later tested each hemisphere's memory for the scenes. During the recognition test, the left hemisphere falsely remembered scenes that were not present in the study phase, but that fit the gist of the presented scenes. The right hemisphere, on the other hand, demonstrated a veridical memory for the scenes. Again, the results suggest that the left hemisphere created inferences and bridged semantic gaps, which led to the false recognition of items that fit a schema or gist.

Research on split-brain subjects clearly establishes the left hemisphere's role in inference making, but only moderately addresses the evaluative capabilities of the right hemisphere. However, there is some evidence from split-brain studies that suggests the right hemisphere is necessary for evaluating and updating interpretations. Miller et al. (2010) presented moral reasoning problems to split-brain patients. The moral reasoning problems consisted of scenarios that featured characters with either cruel or helpful intentions and outcomes that were either harmful or neutral. Their results are consistent with the idea that the split-brain patients' left hemispheres formed quick-and-dirty inferences when presented with initial evidence about the scenario (the potential danger of the situation), and these inferences were resistant to revision when additional evidence about the characters' intentions was provided. This suggests that, without the right hemisphere's ability to reject inappropriate or improbable hypotheses when contradictory evidence arises, the left hemisphere maintains its original explanations, even if they are no longer relevant.

A consistent story emerges from split-brain studies: the left hemisphere is specialized for making inferences, creating explanations, interpreting ambiguity, and bridging information gaps. As Gazzaniga writes, "Ah, lack of knowledge is of no importance, the left brain will find a solution! Order must be made. The first makes-sense explanation will do." (Gazzniga, 2008). It does not seem to be the case that the right hemisphere has the same inferential abilities as the left hemisphere, but they simply go undetected in split-brain patients. In carefully controlled experiments that probe the inferential abilities of each hemisphere irrespective of language abilities, the left hemisphere makes more inferences than the right hemisphere. This may improve performance in some cases, such as when inferring causality, but hinder performance in other cases, such as when remembering whether an item was encountered before or when predicting random events.

2. Reasoning in the damaged brain

Research on patients with unilateral brain damage also indicates that the left hemisphere creates explanatory inferences and the right hemisphere monitors the plausibility of those inferences in relation to other evidence. A study by Deglin and Kinsbourne (1996) offers a striking example of these divergent hemispheric specialties. In their experiment, patients undergoing electroconvulsive therapy (ECT) were shown two premises with familiar or unfamiliar content and were asked if a given conclusion supported or contradicted the premises. For example, patients were given the premises "Every state has a flag. Zambia is a state" and were asked "Does Zambia have a flag, or not?" Each patient completed the task on three separate occasions: once after ECT was used to suppress activity in the left hemisphere, once after ECT was used to suppress activity in the right hemisphere, and once before receiving ECT (which served as a control condition). Incredibly, even though the same syllogisms were presented to the same patients, the patients' responses differed wildly depending on which hemisphere was suppressed. When patients' right hemispheres were suppressed (and their left hemispheres were spared), their responses were consistent with the logic of the syllogism -- even when the logic conflicted the patients' beliefs and when the material was unfamiliar – and the patients responded faster and with more confidence, as compared to the control condition. In contrast, when the left hemisphere was suppressed (and the right hemisphere was spared) the patients tended to give responses that were consistent with real-world prior knowledge. Furthermore, if the premises contradicted real-world knowledge or were unfamiliar, the patients questioned the premises, answered with uncertainty, or refused to answer at all. For example, when presented with the Zambia syllogism, the same subject responded, "Each state has a flag, Zambia has also" under right hemisphere suppression but "Who knows it, this Zambia, how can I know whether it has a flag or not?" under left hemisphere suppression. This research demonstrates the different cognitive tendencies of the hemispheres: the left hemisphere readily makes inferences, regardless of content or conflicting evidence, and the right hemisphere ensures that conclusions are consistent with reality.

Studies on brain-damaged patients are consistent with these results. Patients with an intact left hemisphere, but damaged right hemisphere, retain the ability to make simple inferences about hypothetical situations (Caplan and Dapretto, 2001; Ferstl, Guthke, & von Cramon, 2002; Goel et al., 2006; Reverberi, Shallice, D'Agostini, Skrap, & Bonatti, 2009). Indeed, right hemisphere brain damage is often associated with excessive inference making – patients draw conclusions with incomplete information and infer relationships where there are none (Goel et al., 2006). Patients with unilateral damage to the left hemisphere, on the other hand, have trouble making inferences

(Reverberi et al., 2009). Together, these results demonstrate that causal inference making is lateralized to the left hemisphere to some extent.

Research on unilateral brain damaged patients also suggests that the right hemisphere is specialized for 1) detecting inconsistencies between hypotheses and reality and 2) overturning inappropriate hypotheses. Patients with an intact right hemisphere, but damaged left hemisphere, are sensitive to conflict and can easily identify semantic inconsistencies (Ferstl et al., 2002). When the right hemisphere is damaged, however, patients fail to detect contradictions and their hypotheses become rigid and impermeable to conflicting evidence. In one experiment, Brownell Potter, Bihrle, & Gardner (1986) showed patients with unilateral right hemisphere damage two sentences that could be integrated to form one interpretation. Importantly, one of the sentences supported a different, incorrect interpretation when presented by itself. Brownell manipulated the position of the misleading sentence to test the patients' ability to revise an initial incorrect hypothesis. Compared to controls, patients with right hemisphere brain damage made fewer correct inferences. They could accurately extract true causal relationships but tended to make inappropriate associations between non-related items. Patients' erroneous inferences were especially prevalent when the misleading information was presented first. These findings suggest that an isolated left hemisphere readily creates explanations, but fails to revise them in accordance with new evidence. A functioning right hemisphere is needed to detect inconsistencies and update flawed hypotheses accordingly.

In another study (Danckert, Stottinger, Quehl, & Anderson, 2011), patients with left brain damage, right brain damage, or no brain damage played a game of rock, paper, scissors against a computer. The virtual opponent initially started off with equal choices of rock, paper, and scissors, but eventually changed its responding to heavily favor one of the options. Patients with left brain damage and control participants detected the changes in the virtual opponent's gameplay and updated their strategy accordingly to win more games. Patients with right hemisphere brain damage, however, failed to update their strategy. Again, this suggests that the right hemisphere is necessary to update hypotheses and behaviors to be consistent with new evidence.

In a similar vein, right hemisphere brain damage and hypometabolism have been linked to greater perseveration in the Wisconsin Card Sorting Test (Lombardi et al., 1999). Patients can discover the correct rule but, without a functioning right hemisphere, their beliefs become fixed and immune to subsequent negative feedback. However, it should be noted that other studies that examine performance on the WCST find perseveration in patient groups with both right and left hemisphere damage (Stuss et al., 2000).

The importance of the right hemisphere in shifting from one mental set to another is also apparent in real-world tasks. Goel et al. (2013) asked patients with unilateral brain lesions to plan a trip to Italy and found that patients with right hemisphere brain damage made inferior plans compared with other patients and controls. Goel et al. attribute the patients' poor planning to substandard mental set shifting: "Damage to the right PFC system impairs the encoding and processing of more abstract and vague representations that facilitate lateral transformations. This results in prematurely locking onto precise concrete patterns, and quickly drawing conclusions, albeit substandard ones" (Goel et al., 2013 pg. 721). Without a functioning right prefrontal cortex, patients latched on to potential solutions too quickly and prematurely stopped looking for, considering, and elaborating on alternative solutions.

Together, these studies suggest that the left hemisphere is specialized for creating inferences and the right hemisphere is specialized for monitoring these inferences and overturning them if necessary. In line with this framework, left hemisphere brain damage leads to impaired inference making and a greater reliance on prior knowledge, but conflict detection is spared. In contrast, right hemisphere brain damage does not impair inference making, but patients' inferences can become inappropriately excessive and resistant to contradictory evidence.

3. Reasoning in the deluded brain

Much like how perceptual researchers investigate the normal functioning of the visual system by using optical illusions to study its failures, we can gain insights about normal hypothesis-making by studying cases of abnormal reasoning, such as delusions. Delusions are defined as "fixed beliefs that are not amenable to change in light of conflicting evidence" (American Psychiatric Association, 2013). They are characterized by excessive inference making (Braun & Suffren, 2011), the tendency to prematurely jump to conclusions (Conway et al., 2002; Dudley, John, Young, & Over, 1997; Huq, Garety, Hemsley, & Park, 1988; Moritz & Woodward, 2005; Warman, Lysaker, Martin, Davis, & Haudenschield, 2007), impaired belief updating (Coltheart, Langdon, & McKay, 2011; Coltheart, Menzies, & Sutton, 2010; Coltheart, 2010; Marcel, Tegnér, & Nimmo-Smith, 2004), and the discounting of evidence that disconfirms their beliefs (Woodward, Moritz, Cuttler, & Whitman, 2006; Speechley, Ngan, Moritz, & Woodward, 2010, Eisenacher & Zink, 2017). Based on our predictions that the left hemisphere supports explanatory inference making and the right hemisphere supports belief monitoring and revision, it follows that delusions should be associated with left hemisphere hyperactivity and right hemisphere hypoactivity or damage.

Consistent with this prediction, delusional disorders are strongly linked to damage to the right hemisphere. Braun and Suffren (2011) conducted a meta-analysis of unilateral brain lesions that resulted in delusional disorder and found that 80% of the lesions were right lateralized. Right-lateralized lesions were associated with other delusional symptoms, including somatoparaphrenia (the pathological denial of limb ownership), reduplicative paramnesia (the belief that a place has been duplicated), and flight of ideas. A more recent review of 61 patients with delusional misidentification syndrome, found that 92% of the patients had right hemisphere damage and 63% of patients had damage to the right frontal lobe in particular (Darby & Prasad, 2016). Similarly, Devinsky (2009) reviewed the etiology of delusions and found that several different types of delusions are associated with unilateral right hemisphere or bilateral brain damage, but rarely unilateral left hemisphere damage. Without the constraint of an evaluative right hemisphere, the left hemisphere liberally creates inferences and explanations that are resistant to contradictory evidence and may develop into pathological delusional beliefs. Indeed, as Devinsky writes, "Delusions result from right hemisphere lesions. But it is the left hemisphere that is deluded" (pg. 85). In another review of the role of the right hemisphere in delusions, Gurin and Blum (2017) conclude, the "right hemisphere is essential for our ability to create and maintain accurate appraisals of mental objects holistically and in context."

Corlett et al. (2007) provide additional evidence that delusions arise from a dysfunctional right-lateralized evaluative system. In their study, they taught delusional patients and controls associative mappings between foods and allergic outcomes. After the participants learned the mappings thoroughly, the researchers gave participants negative feedback that contradicted the mappings they were taught. In healthy controls, the right ventrolateral prefrontal cortex (vlPFC) demonstrated a reward-prediction error: relative to baseline, its activity increased in response to unexpected feedback and decreased in response to expected feedback. The right vlPFC of delusional patients, however, failed to distinguish between expected and unexpected feedback, suggesting the delusional patients' prediction-error processing was impaired. These results suggest that delusional patients' right hemispheres fail to detect discrepancies between evidence and beliefs, and this impairment may be present without apparent brain damage.

Coltheart also argues that damage to the right lateral prefrontal cortex underlies the formation of delusional beliefs (Coltheart, 2010). In his two-factor account of delusions, he proposes that delusions require 1) an impairment that elicits some sort of explanation and 2) a dysfunctional belief evaluation system that fails to reject implausible explanations for the initial impairment (Coltheart et al., 2011, 2010; Coltheart, 2010; McKay, Langdon, & Coltheart, 2007). According to Coltheart, the initial impairment determines the content of the delusion and varies from case to case. However, he suggests an impaired belief system is universal among delusional patients and results from damage to the right lateral prefrontal cortex. Coltheart et al. (2010) reviewed cases of patients with neurological symptoms that mirror those of delusional patients, but who nonetheless fail to develop delusional beliefs. They proposed that these patients, unlike delusional patients, have an intact right frontal lobe, which allows them to reject implausible hypotheses and update their beliefs in accordance with reality. In a particularly striking example in support of this view, Bisiach et al. (1991) temporarily increased activity in the right hemisphere of a patient who believed her left arm belonged to her mother by irrigating her left ear with cold water. Incredibly, the irrigation abolished the patient's delusion: up to 2 h after the procedure, she rightfully claimed her arm as her own. This finding has since been replicated (Ramachandran, 1996) and supports the necessity of the right hemisphere in updating inappropriate beliefs.

Anosognosia, the pathological denial of impairment or disease, is another example of a delusion that is rooted in right hemisphere brain damage (Ramachandran, 1996; Stone, Halligan, & Greenwood, 1993). Marcel, Tegner, and Nimmo-Smith (2004) examined the beliefs of 64 patients with anosognosia for hemiplegia (denial of paralysis). Of the 64 patients, 44 had unilateral damage to the right hemisphere and 22 had unilateral damage to the left hemisphere. They found that patients with right hemisphere lesions, but not so much those with left hemisphere lesions, consistently overestimated their abilities to perform tasks requiring the use of their paralyzed limbs. Moreover, only patients with right hemisphere brain damage continued to overestimate their ability to perform a task after they had just failed to perform the task in question. These results illustrate that right hemisphere brain damage cannot only lead to the formation of delusional beliefs, but it can also lead to the persistence of these delusions since inappropriate beliefs are not updated in light of contradictory evidence.

In addition to right hemisphere hypoactivity or damage, delusions may also arise from left hemisphere hyperactivity. Braun and Suffren (2011) review several instances in which delusions arise from left hemisphere hyper-metabolism in the absence of right hemisphere damage. They also note that antipsychotic drugs may reduce delusional ideation by suppressing activity in the left hemisphere and facilitating activity in the right hemisphere. In a similar vein, Mucci et al. (2005) found a relationship between paranoid beliefs in healthy subjects and greater left-lateralized brain activity.

Our framework predicts that normal reasoning goes awry when the balance between left hemisphere inference making and right hemisphere evaluation and revision is disrupted. Delusions are examples of such a disruption. Delusions are characterized by liberal hypothesis formation and impaired hypothesis evaluation and, as would be predicted, are tied to left hemisphere over-activity or right hemisphere under-activity.

4. Reasoning in the healthy brain

The closest correlates to patient studies in non-patient populations are neurostimulation studies. Neurostimulation techniques temporarily alter brain activity and allow researchers to examine corresponding changes in behavior. There have been relatively few neurostimulation studies on reasoning, but several of the studies that have been carried are consistent with the patient literature. Sharot et al. (2012) investigated how inhibition of the left inferior frontal gyrus (IFG) or right IFG with transcranial magnetic stimulation (TMS) influenced how participants incorporated bad news and good news into their beliefs. Individuals generally incorporate good news into their beliefs to a greater extent that bad news (for example, learning that a disease is less common than originally thought will make participants lower their predictions of the prevalence of the disease to a larger than extent than they would raise their prevalence predictions after learning that a disease is more common than originally thought). They found that inhibition of the left IFG, but not inhibition of the right IFG, reduced the good news / bad news effect, meaning that participants were more likely to incorporate negative evidence into their beliefs when the left IFG was inhibited and the right IFG was spared. Lupyan, Mirman, Hamilton, and Thompson-Schill (2012) applied either cathodal tDCS or anodal tDCS to participants' left IFG as they categorized items along different dimensions that ranged from more concrete (for example, color) to more abstract (for example, objects that hold water). They found that participants lowered their thresholds for selecting objects that were more weakly associated with the target category under anodal left IFG stimulation, but not under cathodal stimulation. This finding reinforces the idea that the left hemisphere is driven toward making conceptual associations and inferences. Coltheart et al. (2018) found that inhibiting the right dorsolateral prefrontal cortex (dlPFC) with TMS made healthy participants more susceptible to hypnosis (that is, more susceptible to having "magnetic hands," "levitating arms," "rigid arms," and "sour tastes"), suggesting that the right frontal lobe helps prevent the formation of implausible beliefs and delusions. Finally, several neurostimulation studies have found that the right IFG plays a role in behavioral inhibition (Cunillera, Fuentemilla, Briganani, Cucurell, & Miniussi, 2014; Ditye, Jacobson, Walsh, & Lavidor, 2012; Drummond, Cressman, & Carlsen, 2017; Hogeveen et al., 2016; Jacobson, Javitt, & Lavidor, 2011; Stramaccia et al., 2015).

Using behavioral and neuroimaging methods to study hemispheric lateralization in healthy brains poses several challenges. First, unlike research on brain-damaged and split-brain patients, research on healthy individuals cannot rely on participants' behavior to elucidate brain function. In patient studies, patients' reasoning errors can be analyzed in conjunction with their pathology to determine the function of different brain regions. However, healthy participants' behavior and reasoning errors cannot be localized to one hemisphere or the other, since both hemispheres can contribute to behavior and cognition in an intact brain. Neuroimaging techniques are needed to identify the brain regions that are recruited during reasoning. However, these techniques may still fail to shed light on hemispheric lateralization. Healthy participants naturally form, evaluate, and revise hypotheses concurrently, which complicates the identification of brain networks that support each individual sub-process, if they exist. In order to identify the brain areas that support inference making or belief updating, researchers must use tasks that differentially load on either process. Even then, contrasts designed to find the neural correlates of one process may be contaminated by the neural fingerprints of other processes, which may in turn make a truly lateralized process appear bilateral. Despite these challenges, there is some evidence of a lateralized hypothesis-making network in healthy individuals.

Neuroimaging studies with tasks that feature interpretation and inference making elicit largely left-lateralized brain activity. In one experiment, Parris et al. (2009) recorded subjects' brain activity with fMRI as they watched videos of magic tricks or control videos, which had the same beginning as the magic trick videos but ended with either an expected or unrelated surprising event. They found that only the left dorsolateral PFC was more active when subjects viewed magic tricks than when they viewed expected or unrelated surprising events. As the researchers concluded, this increased activity may reflect the left hemisphere's attempt to explain the magic trick. Left-lateralized brain networks have been identified for tasks involving interpreting and reasoning about past events (D'Argembeau et al., 2013), making inferences about related sentences (Ferstl & von Cramon, 2002; Friese, Rutschmann, Raabe, & Schmalhofer, 2008), and inferring a rule or relationship (Rodriguez-Moreno & Hirsch, 2009).

Tasks involving evaluation and belief updating, on the other hand, implicate the right hemisphere. Neuroimaging studies using healthy subjects consistently find that the right lateral PFC is recruited when logic conflicts with prior beliefs (Goel, Buchel, Frith & Dolan, 2000; Goel & Dolan, 2003; Goel, 2007; Menenti et al., 2009; Stollstorff, Vartanian, & Goel, 2012) and its activity is modulated by the degree to which reasoning problems conflict with real-world knowledge (Stollstorff et al., 2012). Regions in the right hemisphere are also active when subjects receive negative feedback indicating that their hypothesis is no longer correct (Konishi et al., 2002).

The right hemisphere has been shown to support mental set shifts. Cools, Clark, Owen, and Robbins (2002) showed that the right vlPFC is active during the last trial before a hypothesis switch in the Wisconsin Card Sorting Test (WCST) and this activity is independent of the receipt of negative feedback. In the realm of hypothesis making, the right vlPFC may reject a previously held belief or hypothesis and initiate a search for a new one, thus shifting the mental set from one hypothesis to another. Vartanian and Goel (2005) also highlight the role of the right vlPFC in set-shifting. They gave participants problems that required the participants to undergo lateral transformations – mental movements "from one state in a problem space to a horizontally displaced state rather than a more detailed version of the same state" – to correctly solve the problems (Vartanian & Goel, 2005, pg. 1170). Based on the neuroimaging results and previous research, they concluded the right vlPFC initiates lateral transformations. The right vlPFC's involvement in set-shifting and the detection of belief-logic conflicts strengthens the view that it supports belief evaluation and revision. In healthy indiviudals, the right vlPFC may monitor the consistency and plausibility of hypotheses and initiate the revision or search for a new hypothesis if new, contradictory evidence arises.

Outside the domains of reasoning and hypothesis making, the right hemisphere has been shown to play a general role in monitoring and inhibiting thoughts and behavior. Chatham et al.

(2012) and Stuss and Alexander (2007) review evidence that suggests that the right hemisphere, and specifically the right vlPFC, plays a role in context monitoring. Aron, Robbins, and Poldrack (2014) propose that the right vlPFC serves as a brake to inhibit task-irrelevant behavior. Together, these general findings suggest that the right hemisphere monitors the environment for inconsistencies or goal-irrelevant stimuli and inhibits behaviors or thoughts that interfere with the goal at hand.

Although several neuroimaging studies provide evidence for hemispheric asymmetry in reasoning, others report bilateral activations for contrasts that we would predict to have a lateralization component, such as contrasts for inference making (Fangmeier, Knauff, Ruff, & Sloutsky, 2006; Kalbfleisch, Van Meter, & Zeffiro, 2007; Landmann et al., 2006; Lie, Specht, Marshall, & Fink, 2006; Simard et al., 2011) or hypothesis evaluation (Fangmeier et al., 2006; Kroger, Nystrom, Cohen, & Johnson-Laird, 2008; Rodriguez-Moreno & Hirsch, 2009). As discussed earlier, it is possible that these contrasts did not find lateralized processing because the contrasts did not adequately isolate the reasoning process of interest. Indeed, none of the neuroimaging contrasts reported in any of these neuroimaging experiments were specifically designed to test hypotheses about lateralization. In order to gain a greater understanding of the hemispheric asymmetry in reasoning in the neuroimaging literature, we previously ran a meta-analysis of all fMRI studies that involved hypothesis formation or hypothesis evaluation and computed lateralization scores in different brain regions.

Meta-analysis of fMRI studies investigating the neural basis of reasoning: We collected relevant neuroimaging articles by running a series of PubMed searches on 17 terms related to reasoning (e.g., “syllogistic reasoning”, “Raven’s Progressive Matrices”, etc.) and excluded studies that included only patients, used a methodology other than fMRI, or were not published in peer-reviewed journals, resulting in 125 papers (Turner, Marinsek, Ryhal, & Miller, 2015). Two readers independently labeled every contrast in each paper based on the task used and the fMRI analyses implemented. We conducted a meta-analysis of several labels using the Activation Likelihood Estimation approach (Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012) and GingerALE (Eickhoff et al., 2009). The selected labels were chosen to load primarily on hypothesis formation (including "building a model" and "rule finding"), hypothesis evaluation (including "statement verification" and "rule checking"), or conflict. We predicted that the hypothesis formation contrasts would be more left-lateralized and the hypothesis evaluation and conflict contrasts would be more right-lateralized.

The maps shown in Figure 1 present the results of the five separate ALE analyses. In line with the patient literature, conflict is strongly, although not exclusively, right-lateralized and hypothesis formation (consisting of the "building a model" and "rule finding" labels) is predominantly, but again not entirely, left-lateralized. Hypothesis evaluation (consisting of the "statement verification" and "rule checking" labels) shows essentially the same pattern of left-hemisphere dominance as hypothesis formation, which is inconsistent with both the patient literature and our predictions.

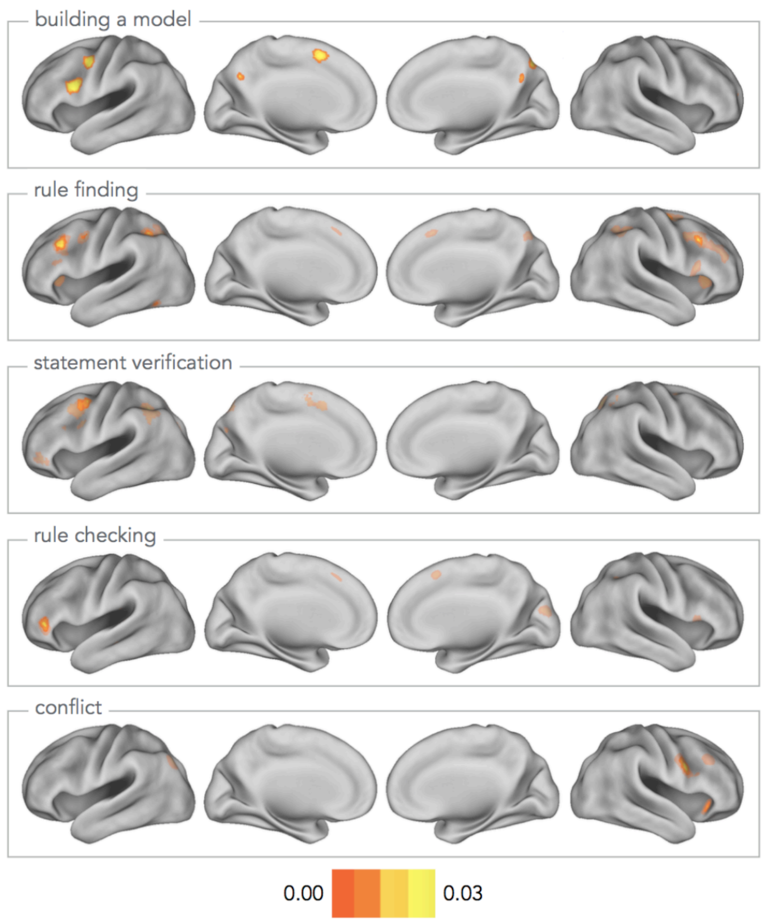

Figure 1: Results of our meta-analyses. The "building a model" and "rule finding" contrasts were predicted to be more left-lateralized and the "statement verification," "rule checking," and "conflict" contrasts were predicted to be more right-lateralized. Translucent clusters correspond to type II error–matched thresholding; solid clusters correspond to type I error–matched thresholding.

There are a number of potential explanations for these ambiguous results. First, it is possible that reasoning has no lateralization components and the laterality observed in patient groups does not represent the true state of the brain. It is also possible that the right hemisphere is specialized for detecting and rejecting salient conflicts between hypotheses and reality, but the hemispheres contribute equally to making simple verifications and evaluations (or these evaluations are done in a content-specific manner). Since many of the contrasts that contributed to the "statement verification" and "rule checking" labels were simple and abstract, these labels may not have been a good test of our hypothesis. Alternatively, it is possible that previous neuroimaging studies fail to capture true lateralization in reasoning. Most of the studies included in our meta-analysis were not designed to separate out the processes of interest and many included confounded processes, in particular those related to language processing or spatial reasoning. It is also possible that fMRI is fundamentally ill-suited to address questions of laterality; for example, different neural architectures may produce different outputs, thus leading to different behaviors, but have similar BOLD activity profiles. These difficulties, even under the "best case" where power issues are minimized, indicate the need for a different approach. This proposal describes three experiments that use a combination of neuroimaging and neurostimulation methods to causally test the hypothesis that the hemispheres have different reasoning biases.

B. Proposed framework of hemispheric asymmetry in reasoning

Based on the evidence presented above, we propose that the left and right hemispheres play different, yet complementary, roles in inferential reasoning. Evidence suggests that the left hemisphere tends to create inferences and explanations to reduce uncertainty. As Gazzaniga suggested nearly three decades ago, the left hemisphere is an interpreter (Gazzaniga, 1989). Its propensity to explain resembles abductive inference, or inference to the best explanation (Coltheart et al., 2010); its inferences do not necessarily have to be correct, or even plausible in some cases, as long as they bridge gaps in information and create a cohesive story. When free from the reign of the moderating influence of the right hemisphere, the left hemisphere's inferences may become excessive or inappropriate, and become resistant to contradictory evidence. However, this is not to say that the left hemisphere always makes poor explanations. Indeed, it seems as if the quality of the left hemisphere's inferences is commensurate to the quality of the evidence it reasons with; most of the time, the left hemisphere's inferences are sound and effectively reduce the uncertainty of the environment. If evidence is sparse or unusual however, the left hemisphere may create explanations that are incorrect, implausible, or even bizarre.

Unlike the left hemisphere, the right hemisphere, in our view, places a premium on cohesion. It is sensitive to conflicts between hypotheses and real-world knowledge or evidence. In line with its roles in belief updating and set-shifting, we propose that the right hemisphere 1) monitors the plausibility of hypotheses and 2) jettisons explanations that are inconsistent with reality or new evidence. Thus, the right hemisphere may prompt the revision of inappropriate hypotheses and initiate the search for new ones. In the realm of hypothesis making, if the left hemisphere is considered an interpreter, the right hemisphere may be considered a realist. While the left hemisphere strives to reduce uncertainty, the right hemisphere strives to reduce inconsistencies between hypotheses and reality.

Our conceptualizations of the left hemisphere as an interpreter and the right hemisphere as a realist are akin to Ramachandran's distinction between left and right hemispheric tendencies (Ramachandran, 1996). According to Ramachandran, "The left hemisphere's job is to create a model and maintain it at all costs... The right hemisphere's strategy, on the other hand, is fundamentally different. I like to call it the 'anomaly detector', for when the anomalous information reaches a certain threshold, the right hemisphere decides that it is time to force the left hemisphere to revise the entire model and start from scratch" (pg. 351-352). Once the left hemisphere adopts a model that minimizes uncertainty and maximizes explanatory power, it incorporates new evidence into the model and may rationalize, ignore, or deny any contradictory evidence. This "band-aid approach" creates a patchwork of explanations and rationalizations, but effectively reduces uncertainty by doing so. Unlike the left hemisphere, the neural processes clustered in the right hemisphere are sensitive to contradictions between beliefs and evidence. When contradictory evidence arises, the right hemisphere may prompt the reworking of the model to satisfy the new evidence or may suppress attempts at explanation altogether – in its view, it is better to be uncertain than wrong.

It's important to note that, although we discuss each hemisphere as a whole, we ascribe different reasoning biases to different local neural networks in each hemisphere. We suggest that differences in neural architecture or functional connectivity can give rise to different processing characteristics and strategies, but these differences likely do not extend to all networks in each hemisphere. When we refer to each hemisphere, rather than the neural networks that comprise it, it is only for the sake of brevity.

1. Congruence with other theories

Our framework is compatible with broader, more general theories of hemispheric lateralization. Bowden, Jung-Beeman, Fleck, and Kounios (2005) propose that the hemispheres differ in the granularity of their computations and these differences produce distinct cognitive strategies and capabilities. Specifically, they suggest that computations in the right hemisphere are more coarsely tuned than those of the left hemisphere, an idea supported by the finding that the right hemisphere is more interconnected—both on a cellular and systems level—than the left hemisphere (Bowden et al., 2005). The relative coarse coding of the right hemisphere and relative fine coding of the left hemisphere would make the hemispheres predisposed to global and local processing, respectively. In the realm of inferential reasoning, the coarse, global processing of the right hemisphere may facilitate the detection of discrepancies between an explanation and its global, real-world context. Conversely, the finer coding of the left hemisphere may emphasize local cohesion at the expense of global consistency.

Braun's "psychic tonus" model of hemispheric specialization is also consistent with our claim that the left hemisphere tends to create inferences while the right hemisphere tends to monitor, evaluate, and revise them (Braun, 2007). According to his model, the left hemisphere generally activates mentation and behavior while the right hemisphere inhibits it. More colloquially, the left hemisphere is predisposed to “do something” and the right hemisphere is inclined to “freeze and recoup” (Braun, 2007, pg. 418). The propensity of the left hemisphere to act and create is consistent with our view that it plays a prominent role in making inferences and bridging gaps in information. Likewise, the right hemisphere’s tendency to inhibit behavior is in line with our proposal that it plays a role in monitoring and inhibiting inappropriate hypotheses.

In his review, Braun also suggests that the hemispheres act in opposition to each other: the suppression of cognitive modules in one hemisphere activates their counterparts in the other hemisphere and vice versa. Craig (2005) echoes the idea of hemispheric opposition in his review of emotional asymmetry and suggests that the balance between opposing hemispheric systems facilitates homeostasis and health. It is possible that hemispheric opposition also plays a role in inferential reasoning; the two hemispheric reasoning systems may compete during inference making and may be preferentially recruited in accordance with situational demands. A dual reasoning system with opponent interactions could promote balanced and flexible inference making. Moreover, disruption in the balance between the two component systems may explain aberrant reasoning in brain-damaged and delusional patients. In the case of right hemisphere brain damage, an overactive left–lateralized reasoning system may lead to excessive inference making. Conversely, a right-lateralized reasoning system resulting from left hemisphere brain damage may lead to insufficient inference making, as patient studies suggest.

In his theory of hemispheric lateralization, Corballis (1989) proposes that the left hemisphere is uniquely generative; that is, it combines visual, lexical, or semantic elements to create “novel assemblages” (pg. 499). Corballis argues that the left hemisphere’s capacity for language and visual image production is rooted in its ability to create novel combinations. In the realm of reasoning, the generativity of the left hemisphere could facilitate, or perhaps even underlie, inference making, since both processes involve making new associations to create a cohesive whole. Unlike the left hemisphere, Corballis suggests that the right hemisphere uses an analogue code of representation; that is, its neural representations reflect reality and leave less room for interpretation and manipulation. Again, the right hemisphere's tendency to represent items veridically is consistent with our view that the right hemisphere prioritizes truth and consistency during reasoning.

In their review of the functional anatomy of inferential reasoning, Barbey & Patterson (2011) presented evidence that the ventrolateral PFC supports explanation generation, the dorsolateral PFC supports explanation evaluation, and the anterior PFC supports the integration of inferences. Although they did not emphasize hemispheric specialization in the review, their meta-analysis revealed a distinct hemispheric lateralization, such that the left hemisphere supports hypothesis formation and integration and the right hemisphere supports hypothesis evaluation. These findings are consistent with our view that the left hemisphere plays a prominent role in forming inferences and the right hemisphere plays a prominent role in monitoring the validity of inferences.

Finally, Goel (2015) offers a similar view of the differential roles of the left and right frontal lobes in reasoning, but emphasized the role of the right hemisphere in maintaining indeterminacy, rather than in resolving conflict. Indeterminacy is similar to uncertainty, but also applies to situations that do not involve probabilities. For example, if you are told that A > B and A > C, you cannot infer anything about the relationship between B and C, so the relationship is indeterminate. Goel posited that the right hemisphere plays a specialized role in maintaining and enhancing indeterminacy, while the left hemisphere attempts to eliminate indeterminacy at all costs. Maintaining indeterminacy can be especially useful in open-ended problems, in which locking onto an early hypothesis can be detrimental to finding the optimal solution. In particularly strong support of this view, one study (Mayseless & Shamay-Tsoory, 2015) found that simultaneously applying anodal tDCS to the right IFG and cathodal tDCS to the left IFG enhanced participants' performance on a divergent verbal reasoning task. Additionally, there was no effect on reasoning when anodal stimulation was applied to the left IFG and cathodal stimulation was applied to the right IFG, or when only anodal stimulation was applied to the right IFG, or when only cathodal stimulation was applied to the left IFG. A similar study (Luft, Zioga, Banissy, & Bhattacharya, 2017) found that cathodal tDCS to the left dlPFC, but not anodal or sham stimulation to the same region, was associated with solving more matchstick problems, and in particular the problems that required participants to relax their assumptions and "think outside of the box."

2. Limitations, challenges, and open questions

The left hemisphere vs. right hemisphere distinction presented here may not hold in other domains of causal inference making, such as language comprehension or perception. For example, although the left hemisphere is generally superior at making judgments about causal structure, the right hemisphere has been shown to skillfully make perceptual (Corballis, 2003; Miller & Valsangkar-Smyth, 2005; Roser et al., 2005) and possibly social (Wende et al., 2013) causal judgments and inferences. The right hemisphere may also play a prominent role in processes related to inference making, such as solving problems with insight (Bowden & Jung-Beeman, 2003; Kounios & Beeman, 2014) or comprehending natural language (Jung-Beeman, 2005).

According to our claim that the right hemisphere monitors the validity of inferences, a damaged or disconnected right hemisphere should lead to implausible or excessive inference making and possibly even delusions. However, split-brain patients and some patients with unilateral right hemisphere brain damage fail to develop delusional disorders (Coltheart, 2010; Gazzaniga, 2000). Coltheart's two factor theory of delusions may offer an explanation as to why split-brain patients and patients with right hemisphere brain damage do not succumb to disorders in reasoning. As we stated earlier, Coltheart proposes that delusions require 1) an impairment that elicits some sort of explanation and 2) a dysfunctional belief evaluation system that fails to reject implausible explanations for the initial impairment (Coltheart et al., 2011, 2010; Coltheart, 2010; McKay, Langdon, & Coltheart, 2007). We suggest that split-brain patients and unilateral right brain-damaged patients may meet the second requirement but not the first. That is, these patients may have an impaired evaluation system but they fail to develop delusional beliefs because they lack a neurological impairment that warrants an explanation. Right hemisphere damage or isolation may therefore make a patient vulnerable to abnormal reasoning, but it does not necessarily precipitate delusional ideation; in other words, left hemisphere over-activity or right hemisphere under-activity is necessary, but not sufficient, for delusions to form. It is also likely that – although the right hemisphere plays a more prominent role in hypothesis evaluation – the left hemisphere retains some evaluative capabilities, which may prevent delusional beliefs from forming.

Although several lines of evidence support our framework, some do not. Under our framework, the receipt of inconsistent evidence should preferentially activate regions in the right hemisphere, but Fugelsang & Dunbar (2005) found left-lateralized activations in subjects who were presented with evidence that conflicted a theory. Since this study utilized a block design, however, it is possible that it failed to capture the true neural activity associated with the presentation of contradictory evidence; the left lateralized activations may instead reflect inferential reasoning processes other than hypothesis evaluation and belief updating. A study by Vartanian & Goel (2005) also contradicts our predictions. They propose that the right ventrolateral PFC (vlPFC) supports unconstrained hypothesis generation, since they found that activity in the right vlPFC increases as the constraints of an anagram-solving task decrease. An alternative explanation for their results could be that the right vlPFC plays a role in monitoring the problem solving process or inhibiting unfruitful lines of reasoning, both of which would fit within our framework.

The strongest evidence against our proposal stems from two transcranial magnetic stimulation (TMS) studies conducted by Tsujii and colleagues (Tsujii, Masuda, Akiyama, & Watanabe, 2010;

Tsujii, Sakatani, Masuda, Akiyama, & Watanabe, 2011). In both studies, TMS was applied to the left and right inferior frontal cortices of subjects prior to a deduction task that involved belief-logic conflicts. Since we postulate that the right hemisphere favors truth and real-world knowledge over logic, we would expect inhibition of the left vlPFC, but not the right vlPFC, to cause an over-reliance on prior beliefs. Contrary to our predictions, they found that right vlPFC disruption increased the influence of prior beliefs on logic (thus enhancing the belief-bias effect) and left vlPFC disruption eliminated the belief-bias effect. It is possible that the right hemisphere's role in reasoning is not only to monitor, revise, and reject improbable hypotheses, as we have proposed, but also to inhibit information that is irrelevant to reasoning. In this case, disruption of an inhibitory right vlPFC would allow conflicting prior beliefs to bias reasoning, thus enhancing the belief-bias effect as Tsujii and colleagues found. Alternatively, it is possible that the targeted brain regions were not modulated as intended, even though the stimulation parameters used in the studies have been shown to reduce cortical excitability in motor cortex following stimulation (Robertson, Théoret, & Pascual-Leone, 2003). In any case, additional studies using complementary neuromodulation techniques are needed to corroborate and expand on these findings.

Patient studies provide much stronger support for hemispheric asymmetry in reasoning than neuroimaging studies using healthy individuals. There are several possible explanations for the apparent discrepancy between the patient literature and the neuroimaging literature. The first possibility is that lateralization is an artifact of abnormal brain functioning. Lateralization is exaggerated in brain-damaged and split-brain patients because the damaged (or disconnected) brain tissue 1) cannot contribute to reasoning and 2) may release contralateral brain regions from inhibition, thus amplifying the reasoning biases of the contralateral hemisphere (Braun, 2007). So although patients may exhibit lateralized reasoning, it is possible that neuroimaging captures the true state of the system, which is a bilateral, symmetric reasoning network. Alternatively, it is possible that neuroimaging studies obfuscate a truly lateralized reasoning system because contrasts fail to isolate inference making or belief updating adequately. Indeed, many contrasts designed to identify brain regions associated with inference making are confounded with unrelated processes (Fangmeier et al., 2006; Kalbfleisch et al., 2007; Landmann et al., 2007). These contaminated processes may muddle neuroimaging results, making lateralized processes appear bilateral.

C. Primary aims and rationale

We propose that inferential reasoning strategies are subtly lateralized in the brain, such that cognitive processes in the left hemisphere tend to reduce uncertainty and those in the right hemisphere tend to resolve conflict (Marinsek, Turner, Gazzaniga, & Miller, 2014). These divergent reasoning strategies make the left hemisphere prone to create explanations, infer causality, and fill in gaps in information, and the right hemisphere prone to detect inconsistencies between hypotheses and other evidence and reject implausible hypotheses. We argue that, in the healthy brain, the different reasoning strategies of the left and right hemispheres create a flexible and balanced reasoning system. Furthermore, neural networks in the left or right hemispheres can be preferentially recruited in order to bias reasoning to meet task demands (making inferences more liberal or cautious, for example). Our framework is strongly supported by patient studies, including research on brain-damaged patients, delusional patients, and split-brain patients. However, neuroimaging and neurostimulation studies using healthy individuals provide less convincing evidence. To determine whether hemispheric lateralization contributes to healthy human reasoning, or whether it is an artifact of abnormal brain functioning in patients, we designed and carried out a series of neurostimulation and neuroimaging experiments.

We used convergent approaches to test the idea that each hemisphere possesses different reasoning biases. Our experiments were designed with the following aims:

1. Establish causal relationships between reasoning biases and lateralized brain activity by observing how tDCS-induced hemispheric asymmetry influences participants' reasoning behavior on a total of four reasoning tasks, one of which was designed to be amenable to cognitive modeling and three of which were designed to closely resemble real-world, nuanced reasoning scenarios.
2. Identify brain networks that are active during uncertainty (which we predict will be mostly left-lateralized) and networks that are active when evidence conflicts beliefs (which we predict will be mostly right-lateralized) by recording participants' brain activity with functional magnetic resonance imaging (fMRI) as they complete a novel probabilistic inference task.

1. Stimulation design

Given the fact that previous reports of hemispheric asymmetry in other domains have been found to be more apparent than real (as discussed in Miller, Kingstone, & Gazzaniga, 2002), it is crucial to use causal methods to test lateralized processing. Until now, few studies have attempted to establish causal relationships between reasoning and brain activity, and the few studies that have used neuromodulation to study reasoning have yielded mixed results: some studies support the predictions of our framework (Coltheart et al., 2019; Lupyan, Mirman, Hamilton & Thompson-Schill, 2012; Sharot et al. 2012; Xue, Juan, Chang, Lu, & Dong, 2012) , some oppose them (Tsujii, Masuda, Akiyama, & Watanabe, 2010; Tsujii, Sakatani, Masuda, Akiyama, & Watanabe, 2011), and other studies find null results (Hecht, Walsh, & Lavidor, 2010).

Existing transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS) studies may conflict with each other because many rely on between-subjects designs and fail to model the current flow induced by the stimulation set-ups. In order to establish a clear link between brain activity and reasoning, it is imperative to model how current is delivered to all brain regions and on an individual basis, since electrode placement, cortical folding, and skull shunting can drastically (and non-intuitively) alter the effects of stimulation (Miranda, Lomarev, & Hallett, 2006; Nitsche et al., 2008, Datta et al., 2009). In order to circumvent the problems of previous studies, we used a within-subjects design and modeled the current flow induced by high-definition tDCS (HD-tDCS) for every participant. In the tDCS experiments, we temporarily modulated activity in participants' left and right hemispheres and observed their performance in several novel reasoning tasks. In doing so, we aimed to 1) establish causal relationships between brain activity and reasoning behavior, 2) test our hypothesis that differences in hemispheric dominance lead to reasoning biases, and 3) conceptually replicate the results of studies on brain-damaged or delusional patients by comparing their performance to the participants' performance.

High-Definition Transcranial Direct Current Stimulation (HD-tDCS): tDCS is a neuromodulation technique in which a weak electrical current is applied to the scalp in order to temporarily modulate existing brain activity. The intensity of the electrical current used in tDCS protocols is too weak to induce action potentials directly (Nitsche et al., 2008, Stagg & Nitsche, 2012); instead, tDCS is thought to modulate spontaneous neuronal firing by altering neuronal resting membrane potentials (Nitsche & Paulus, 2000). Cathodal stimulation decreases (hyperpolarizes) the membrane potential, thus decreasing excitability and spontaneous firing, and anodal stimulation raises (depolarizes) the resting potential, thus increasing neuronal excitability (Nitsche & Paulus, 2000, Nitsche et al., 2008). tDCS has been shown to increase glutamate release in target brain areas during anodal stimulation (Hone-Blanchet, Edden, & Fecteau, 2016), polarize neurons and modulate responses to neuronal input (Chakraborty, Truong, Bikson, & Kaphzan, 2017), increase global mean field power (Lauro et al., 2014), and modulate cerebral perfusion in target brain areas (Stagg et al., 2013). Although the effects of tDCS are temporary, stimulation periods ranging from 9-13 minutes produce after-effects that last about one hour, and longer stimulation periods produce longer after-effects (Nitsche et al., 2008). An especially attractive feature of tDCS is the fact that it can be incorporated into double-blind, sham-controlled designs. Sham tDCS conditions are identical to active tDCS sessions except that current is only briefly applied to subjects' scalps in order replicate the tingling effects of tDCS without altering brain activity.

Traditional tDCS set-ups use large rectangular sponge electrodes, which modulate current diffusely, stimulate regions outside of the target brain area, and fail to deliver peak current density to brain regions located directly under the sponge anode (Borckardt et al., 2012; Caparelli-Daquer et al., 2012). Several meta-analyses of traditional tDCS studies have found inconsistent results and null effects when results are averaged across studies (Horvath, Forte, & Carter, 2015A, Horvath, Forte, & Carter, 2015B; Jacobson & Koslowsky, 2012; Tremblay et al., 2014). These inconsistent results are likely due to the poor targeting and great variability inherent with sponge-based tDCS systems.

To mitigate the shortcomings of traditional tDCS, we used high-definition tDCS (HD-tDCS). HD-tDCS systems use compact electrodes (much like EEG electrodes) instead of traditional sponge electrodes, which increases the focal specificity of stimulation (Borckardt et al. 2012). HD-tDCS has been shown to modulate cortical excitabtilty (Kuo et al., 2013) and increase blood oxygenation inside but not outside of targeted brain areas (Muthalib, Besson, Rothwell, and Perrey, 2017). We used the HD-Targets modeling software to create an HD-tDCS electrode montage that optimally delievered current to cortical targets. In addition to optimizing electrode placement at a group level, we created anatomical meshes of each participant's brain, adapted the general electrode configuration to their unique cortical folding, modeled current flow in their brain, and used the estimates of cortical current densities as covariates in our behavioral analyses. Furthermore, we used an infrared camera positioning system to place the electrodes precisely across sessions and across participants. These methods allow us to exert a greater level of control over the delivery of current and allow us to accout for individual differences in head size and shape, cerebrospinal fluid volume, fat content, skull thickness, and gyral folding, all which have been shown to influence the amount of current that reaches the cortical surface (Li, Uehara, & Hanakawa, 2015). We are unaware of any tDCS or HD-tDCS study that implemented a more thorough sitmulation protocol.

Stimulation sites: tDCS was used to diffusely boost neural activity in the left or right inferior frontal gyrus (IFG; BA 45) while simultaneously inhibiting neural activity in the contralateral IFG, in order to mimic the hemispheric asymmetry present in patient groups and characterize the effects of hemispheric asymmetry on reasoning. The inferior frontal gyri were chosen as stimulation targets because these regions have been associated with reasoning in the past. Several studies indicate that the right inferior frontal gyrus (IFG) is recruited during hypothesis evaluation (Barbey & Patterson, 2011; Cools, Clark, Owen, & Robbins, 2002; Goel et al., 2000; Goel & Dolan, 2003; Goel, 2007; Menenti et al., 2008; Stollstorff et al., 2012). Conversely, the left IFG is consistently activated during tasks related to hypothesis formation (Barbey & Patterson, 2011; D'Argembeau et al., 2013; Friese et al., 2008; Goel et al., 2000; Monti & Osherson, 2012; Rodriguez-Moreno & Hirsch, 2009).

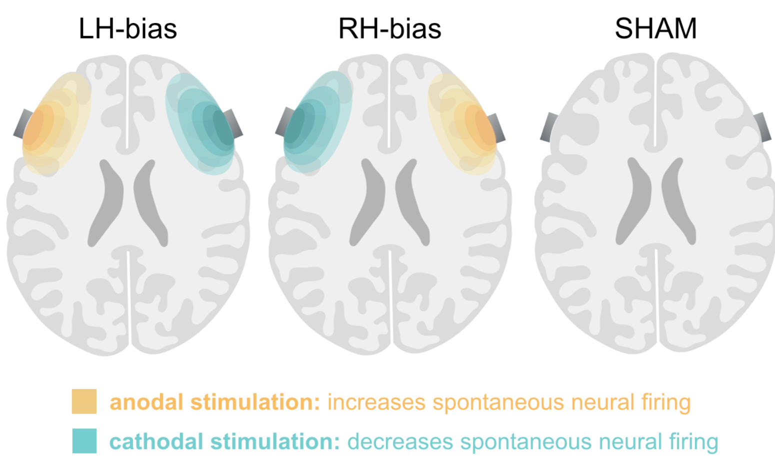

Figure 2: LH-bias, RH-bias, and sham stimulation conditions.

Figure 2 illustrates the general stimulation set-up. Each experiment included three stimulation conditions: a LH-bias stimulation condition, in which anodal HD-tDCS was applied to the left IFG and cathodal HD-tDCS was simultaneously applied to the right IFG; an RH-bias stimulation condition, in which anodal HD-tDCS was applied to the right IFG and cathodal HD-tDCS was simultaneously applied to the left IFG; and a sham condition, in which current was only transiently applied to the bilateral inferior frontal gyri. The LH-bias stimulation condition mimics left hemisphere hyperactivity (which may be comparable to patients with delusions or right hemisphere brain damage) and the RH-bias stimulation condition mimics right hemisphere hyperactivity (which may be comparable to patients with left hemisphere brain damage). The sham stimulation does not induce changes in brain activity, and thus served as a control.

2. Predictions

We predicted that inducing hemispheric asymmetry with HD-tDCS would produce some of the same reasoning biases in healthy individuals as those observed in different patient groups. Specifically, we predicted that LH-bias stimulation would make participants more certain of their beliefs and hypotheses. When participants were uncertain, we predicted that they would update their hypotheses to a greater extent under LH-bias stimulation than under RH-bias stimulation in order to become more certain, and when they were already certain, they would update their hypotheses to a lesser extent in order to maintain high levels of certainty. We predicted that RH-bias stimulation would influence individuals' responses to conflict, such that individuals would become more willing to revise their beliefs when they received conflicting evidence under RH-bias stimulation. We also expected that LH-bias and RH-bias stimulation would exert opposite effects on behavior compared to sham. That is, if LH-bias stimulation increased a particular behavior relative to sham, we expected that RH-bias stimulation would decrease the same behavior relative to sham, and vice versa.

II. Experiment 1: Effects of HD-tDCS on probabilistic inference making

A. Rationale

The goal of this experiment was to temporarily induce hemispheric asymmetry with high-definition transcranial direct current stimulation (HD-tDCS) and examine its effect on different aspects of reasoning, including evidence accumulation, evidence evaluation, belief updating, and decision making. Specifically, this experiment aimed to test the primary predictions of the hemispheric asymmetry framework that 1) left hemisphere networks are driven toward increasing and maintaining certainty and 2) right hemisphere networks are driven toward maintaining cohesion between evidence and beliefs. Testing both of these predictions requires a task in which participants make inferences under uncertainty and accumulate evidence that occasionally conflicts their prior hypotheses. The beads task, a probabilistic inference task that is commonly used to evaluate reasoning in patients with delusions, fulfills both of these requirements. In the beads task, an experimenter shows participants two jars containing different proportions of blue and white beads (for example, jar A contains 70% blue beads and jar B contains 70% white beads). The experimenter hides the jars, selects one of the jars, and then draws a bead from the selected jar and shows the participant the bead. The participants guess which jar was selected based on what beads were drawn and can request to see as many beads as they want. This task requires participants to make probabilistic inferences that are inherently uncertain, evaluate evidence that occasionally conflicts prior hypotheses, establish a threshold for stopping evidence collection, and, importantly, ideal belief updating can be modeled using Bayes theorem. Multiple studies have shown that delusional individuals request fewer beads than non-delusional individuals, who tend to be too conservative (Conway et al., 2002; Dudley, John, Young, & Over, 1997; Huq, Garety, Hemsley, & Park, 1988; Moritz & Woodward, 2005; Warman, Lysaker, Martin, Davis, & Haudenschield, 2007).

The task used in this experiment borrowed the structure of the beads task, but differed in several key ways. First, participants used a digital slider to continuously report their guesses. The digital slider can capture more precise guesses than a Likert scale or two-alternative forced choice and can potentially detect more subtle belief updates. Second, the stimuli were changed to engender more complex probabilistic reasoning and make the task more aligned with real-world decision making. Rather than guessing which jar was chosen based on the color of observed beads, participants in this task guessed which U.S. state was chosen based on the ethnicity of observed residents from the state. Third, participants were not informed about the probabilistic relationships between evidence and hypotheses (as in the beads task where participants are informed, for example, that there's a 30% chance a white ball came from jar A and a 70% chance it came from jar B). Instead, and much like in the real world, this task required participants to rely on their own background knowledge and their implicit associations between evidence and candidate hypotheses.

We predicted that HD-tDCS-induced hemispheric asymmetry would change participants' reasoning behavior in alignment with our hemispheric asymmetry framework and that LH-bias stimulation and RH-bias stimulation would produce changes in opposite directions compared to sham. First, we predicted that LH-bias stimulation would make participants more certain and would materialize in 1) slider positions closer to either end of the track (associated with more certainty), 2) less evidence collection, and 3) a lower threshold for stopping evidence collection. Second, we predicted that LH-bias stimulation would make participants less sensitive to conflicts between beliefs and evidence, which would result in smaller backtracks in strong beliefs after receiving conflicting evidence. We made the opposite predictions for RH-bias stimulation.

B. Methods

1. Participants

26 individuals (18 females, mean age: 21, age range: 18-37) participated in the transcranial direct current stimulation (tDCS) experiment. Participants were recruited by sending emails to several undergraduate mailing lists and by posting a description of the study to an online Psychology participation system. There were strict eligibility criteria that ensured participants could safely undergo MRI and tDCS. Eligibility criteria included 1) no metal in the body, 2) no history of epilepsy, stroke, or brain damage, 3) no pacemaker or brain stimulator, 4) no dredlocks or irritation on the scalp near the electrode sites, and 5) no possibility of pregnancy. The screening form used in Villamar et al. (2013) was used to screen participants. Answering "yes" to any question excluded the individual from participating, with the exception of the question about medications. Only individuals taking psychological medications (such as for depression or anxiety) were excluded.

To be selected to participate in the tDCS study, participants had to pass a behavioral prescreen in which they completed the state guessing task (described below) and a memory task for a separate experiment. A total of 151 participants completed the behavioral prescreen. To be selected to participate in the tDCS experiment, individuals had to meet all eligibility criteria, move the digital slider in at least 50% of the first six evidence presentations, and shift their decision criteria in the memory task. Participants were selected based on their use of the digital slider because many participants failed to move the slider at all before making a decision despite being instructed to move the slider whenever their belief changed. Eligible participants were paid $20 per hour for their participation in an initial MRI session and the three tDCS sessions.

2. Probabilistic inference task

Participants completed a novel probabilistic inference task in which they had to rely on their background knowledge of U.S. state demographics to make inferences. Participants were told that the computer would randomly select one of two possible U.S. states at the beginning of each trial and their goal was to guess which state was selected. After a state was selected, the computer randomly selected one resident from the chosen state and presented the ethnicity of that resident for 3 seconds. Participants were truthfully told that the computer selected residents in accordance with 2010 census data, such that if the chosen state had 72% white residents, the computer would select a white resident with a probability of 0.72, and so on. After a resident's ethnicity was presented, participants used a digital slider with the two possible U.S. state options at either end to report their best guess of which state was selected. Ethnicities continued to be presented every three seconds, and participants were told to update the slider whenever their guess (or belief) changed. After six residents were shown, a stop button appeared and participants were given the option to stop collecting evidence and submit their final guess. A maximum of 20 residents could be observed in every trial. Once participants clicked the stop button or viewed all 20 residents, they were prompted to choose which of the two U.S. states was selected and report how confident they were in their decision on a scale from 1-5, where 1="low", 2="somewhat low", 3="moderate", 4="somewhat high", and 5="high."

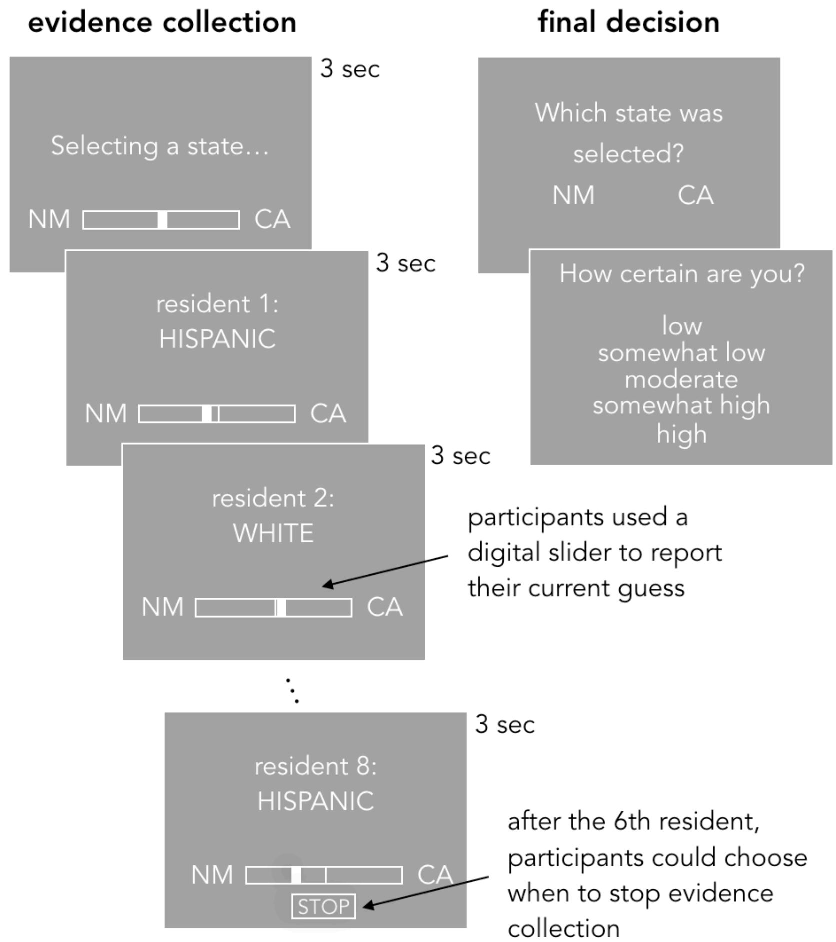

Figure 3: Stimuli presentation in the state guessing task. Participants were presented with ethnicities of residents of a mystery state and had to guess which state was selected. Participants used a digital slider to report their guesses.

In order to encourage participants to use the slider throughout the evidence presentation, cash bonuses were offered based on the slider position. If the slider was in the very center of the track, no money was rewarded or penalized. The closer the slider was to the correct state, the more money participants were rewarded, up to $1.00. The closer the slider was to the incorrect state, the more money was penalized, up to -$1.00. Participants were told that the best way to maximize their bonus was to move the slider toward the state that they thought was correct and make the distance away from the center of the track proportional to their certainty. For every trial in the experiment, one of the evidence presentations was randomly selected and the slider position for that piece of evidence was used to calculate the reward or penalty. The bonus amounts were summed, rounded up to the nearest dollar, and paid in cash at the end of each session.

3. Stimuli creation

Five U.S. states were used in this experiment: California, Iowa, Louisiana, New Mexico, and New York. These states were chosen to so that pairwise matchups between the states would vary in difficulty (for example, California vs. New York is a difficult matchup because the states share similar demographics, while Louisiana vs. New Mexico is an easier matchup). State demographics were pulled from the 2010 census (www.census.gov/2010census/data) and the following races/ethnicities were used: White, Black, Asian, Native American, and Hispanic. The percentages of the White, Black, Asian, and Native American residents in each state were based off of the "Alone" column in Table 4 of The White Population: 2014, Table 5 of The Black Population: 2010, Table 2 of The Asian Population: 2010, and Table 2 of The Native American and Alaska Native Population: 2010 documents, respectively. The percentages of Hispanic residents in the five states in 2010 were taken from Table 2 of The Hispanic Population: 2010. Percentages that did not add up to 100% (due to residents reporting they did not identify as only one of these ethnicities) were normalized so that they added up to 100%.

2500 trials were simulated by 1) randomly selecting two of the five U.S. states without replacement (the first state was designated as the correct state and the second state as the alternative state) and 2) randomly selecting 20 ethnicities with replacement based on proportions of each ethnicity in the correct state. Next, Bayes' theorem was used to compute the probability that the correct state was selected at each observation:

$$P(state_{cor}|eth) = \frac{P(state_{cor}) * P(eth|state_{cor})}{(P(state_{cor}) * P(eth|state_{cor})) + (P(state_{alt}) * P(eth|state_{alt}))}$$

Equation 1: Bayes equation for state-guessing experiment

where $state_{cor}$ refers to the correct U.S. state, $state_{alt}$ refers to the alternate state, and eth refers to the presented resident's ethnicity. The priors for the first resident in each sequence were set to p=0.5, and subsequent priors were set to the posterior for the previous resident.

To ensure that the difficulties of the trials were similar across tDCS sessions, the Bayesian posterior time courses were grouped into clusters with similar trajectories using affinity propagation. Affinity propagation produced 10 clusters. Some of the trajectories within each cluster were not very representative of the mean trajectory of the cluster. These trials were weeded out by computing the sum of the squared residuals between each trajectory and its cluster's mean trajectory and removing trials with a sum of squared residuals greater than 0.75. Trials were randomly selected from 8 of the biggest clusters and assigned to 6 stimuli sets (3 sessions x 2 time points). More trials were pulled from bigger clusters and fewer trials were pulled from smaller clusters. Each stimuli set consisted of 24 trials total.

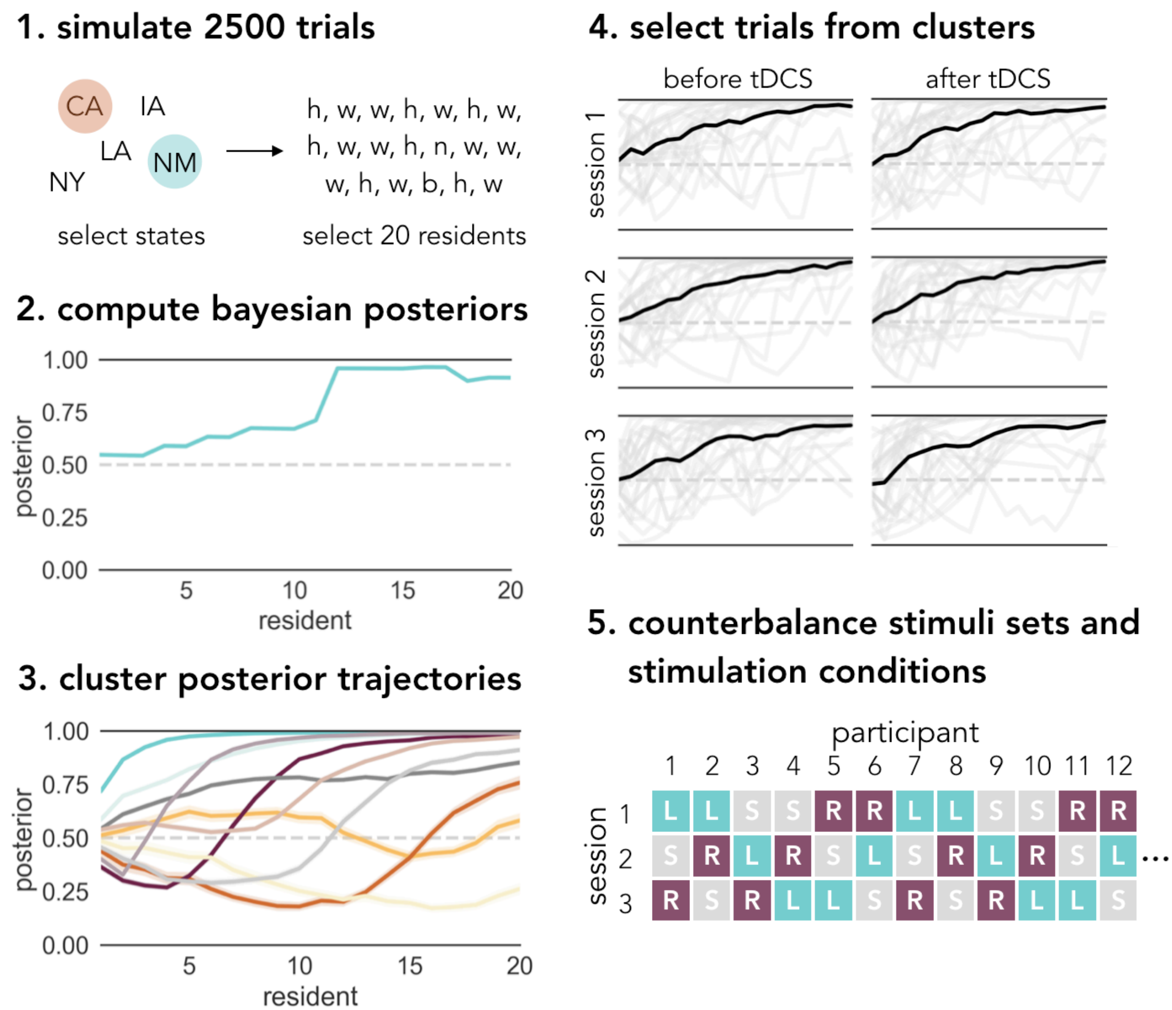

Figure 4: Creation of the stimuli used in the state guessing task.

4. Current flow modeling

The Soterix HD-Targets software was used to determine the optimal electrode placement to deliver maximal current to the left and right inferior frontal gyri. The optimal electrode configuration consisted of only two electrodes, one electrode over the left IFG and one electrode over the right IFG. To account for variations in participants' brain anatomy, we adapted this general electrode configuration for each participant. First, T1-weighted and T2-weighted anatomical scans were obtained for each participant with MRI. Then, the mri2mesh function from SimNIBS (Thielscher, Antunes, & Saturnino, 2015) was used to convert the anatomical volumes into 3D meshes of the skin, skull, cerebrospinal fluid, segmented gray matter, and white matter. The segmented gray matter regions were loaded in Freeview and brain targets were manually defined as the surface RAS coordinates of the center of masses of the left and right inferior frontal gyrus, pars triangularis (BA 45). The center of mass was estimated visually and care was taken to avoid defining a target near a sulcus.

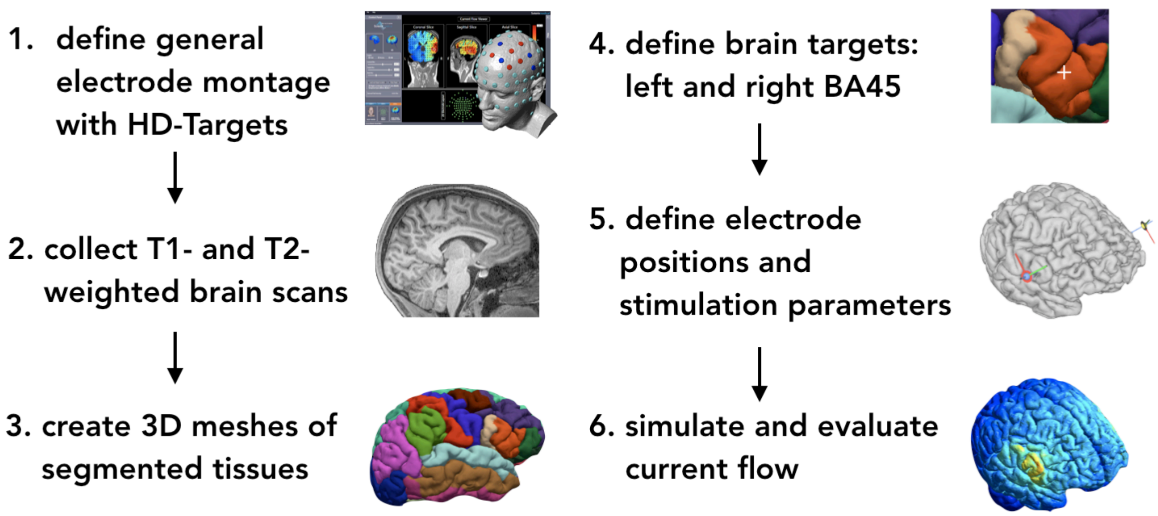

Figure 5: Pipeline for electrode configuration and current modeling.

SimNIBS was used to do define the electrode locations and simulate current flow. The following parameters were used to define the electrodes: current: 2.00mA, shape: elliptical, size: 1.2cm diameter, type: electrode+gel, electrode thickness: 1.6mm, gel thickness: 8.4mm, electrode size ratio: 1.00, connector: whole surface. The surface RAS coordinates were used as the x, y, z coordinates of the electrode. The reference direction was defined as x, y-10, z. The coordinates were projected onto the scalp, and the resulting x,y, and z coordinates were used as the electrode coordinates. Two electrodes were defined, an anode over the left IFG and a cathode over the right IFG, and current flow was simulated. The current modeling results were viewed and validated in gmsh (Geuzaine & Remacle, 2009).

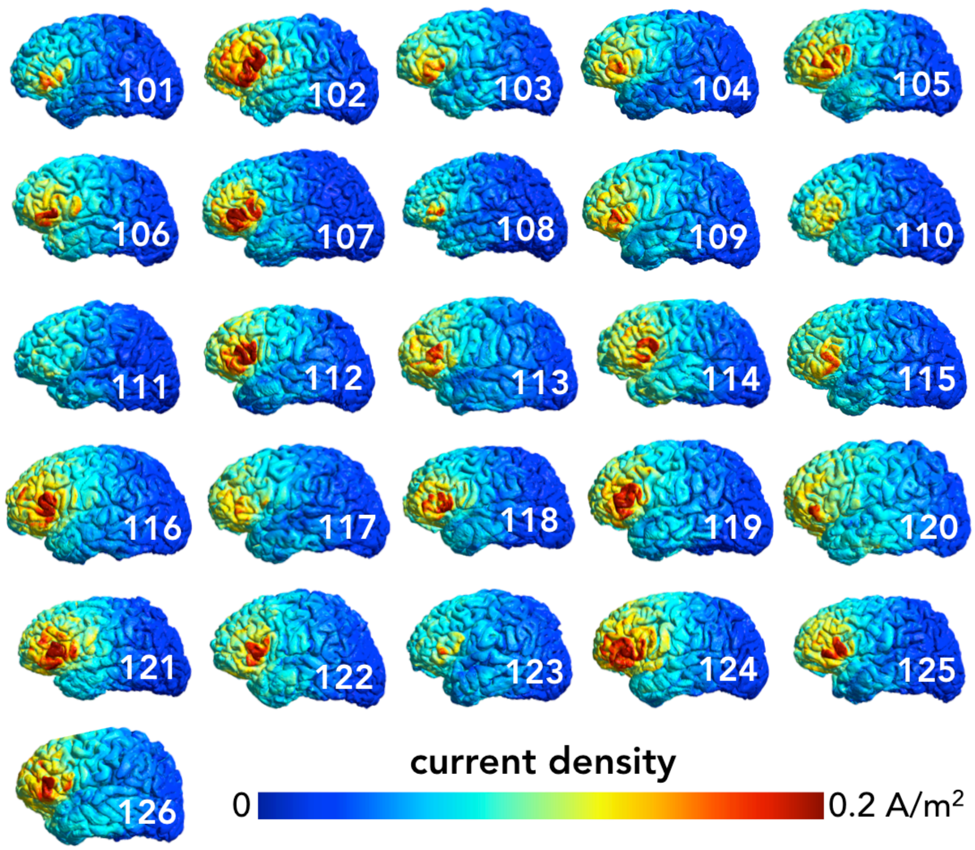

Figure 6: Individual differences in estimated current density at the cortical surface.

5. Procedure

Each participant completed three sessions: a LH-bias condition with a left anode and right cathode, a RH-bias condition with a right anode and left cathode, and a sham condition which served as a control. The stimulation sessions were completed at the same time of day and at least 48 hours apart.

At the beginning of every session, participants completed 24 trials of the task in a behavioral testing room. Then they were fitted with the electrodes. The Brainsight navigation system was used to register the participant's head to their anatomical scan and locate the electrode coordinates on the participant's temples, which were recorded with a marker. An empty electrode cap was used to mark the perimeters of each electrode and the skin within the electrode perimeter was swabbed with alcohol. Gel-filled electrodes (Soterix Medical) were positioned within the electrode perimeters and taped in place. A reference electrode was taped on the forehead to determine the electrode connection qualities. The electrodes were connected to the stimulator: the reference electrode was inserted into channel C, the anode was inserted to channel 1 which was set to +2.0mA, and the cathode was inserted to channel 2 which was set to -2.0mA. If the electrode connection qualities were above 25K, the skin under the electrodes was swabbed with a cotton swab. In rare cases, the connection qualities could not get lower than 25K, so a cutoff of 40K was used. Participants were stimulated for a total of 20 minutes. During stimulation, current slowly ramped up in the first 30 seconds, stayed constant at 2.0mA for 19 minutes, and slowly ramped down in the last 30 seconds. In the sham condition, the current ramped up and then back down in the first 60 seconds, stayed constant at about 0.05mA for 18 minutes, and then ramped up slowly and back down in the last 60 seconds. The electrode setup was identical for the sham session. Odd-numbered participants were given left-anode sham stimulation and even-numbered participants were given right-anode sham stimulation. 30 seconds after the beginning of stimulation, participants were asked to describe the feeling and intensity of the sensations under the electrodes. The task began 90 seconds after the onset of the stimulation. Participants completed 24 trials of the state guessing task. If participants finished the task while the stimulation was still going, they were asked to report the sensations under the electrodes again.

After the third session, participants completed an online survey that consisted of questions about motivation (McAuley, Duncan, & Tammen, 1989), handedness (Oldfield, 1971), political attitudes and belief superiority (Toner, Leary, Asher, & Jongman-Sereno, 2013), dogmatism (Altemeyer, 2002), and delusional ideation (Peters, Joseph, Day, & Garety, 2004). Participants also estimated the percentage of Native American, Asian, Black, Hispanic, and White residents living in California, Iowa, Louisiana, New Mexico, and New York.

6. Data analysis

All analyses were performed in Python using the Pandas, Numpy, Sci-kit learn, and statsmodels packages.

Participants' demographic estimates were used to compute the expected Bayesian posteriors for each observation by substituting participants' estimates for the P(eth|state) terms in Equation 1. In cases where participants' percentages did not sum to 100%, their estimates were scaled so that they summed to 100%.

Multiple linear regression was used to determine the effects of stimulation on 1) the amount of evidence collected and 2) the Bayesian posteriors when evidence collection stopped. The design matrices consisted of the following regressors: block number (0: baseline, 1: during stimulation); the LH session; the RH session; LH-bias (interaction between block * LH session); RH-bias (interaction between block * RH session); LH-bias intensity (demeaned cortical current densities for each participant * LH-bias); RH-bias intensity (demeaned cortical current densities for each participant * RH-bias); degree of right-handedness for each participant (continuous); dummy variables for stimuli sets 1 and 3; trial number; trial difficulty (1 - weighted sum of Bayesian posteriors); dogmatism score for each participant; self-reported effort for each participant; self-reported skill for each participant; and the delusional ideation score for each participant. Regressors were normalized so that they ranged from 0 to 1. Beta values were computed with Ordinary Least Squares and a Bonferroni correction was applied to correct for multiple comparisons ($p<0.003$).

Trajectory regression analysis was used to determine the effect of tDCS on certainty dynamics. Certainty was operationalized as the distance away from the slider midpoint and the certainty at each observation was normalized across trials by subtracting the certainty associated with the ideal Bayesian posterior. Trajectories were defined as the certainty values associated with the first seven residents of each trial. The trajectory regression analysis consisted of clustering certainty trajectories into groups and using multinomial logistic regression to predict group membership. To determine the number of groups to use, k-means clustering was used to cluster the certainty trajectories into 1 to 10 groups. Each group of clusters was scored by summing the distances between each point and the cluster center, and the elbow method was used to select the optimal group size (in this case, k=2). The certainty trajectories were clustered into two groups and logistic regression was used to determine what factors predicted group membership. The same design matrix was used as before.

To determine how prior beliefs and evidence affect participants' belief updates, continuous heatmaps were created that denoted participants' average belief update over the full range of prior belief certainties (from completely uncertain to completely certain) and evidence consistencies (from highly disconfirmatory to highly confirmatory of the prior). The magnitude of the belief updates was equal to the absolute difference in slider position before vs. after each observation and the direction of the update was normalized so that positive values indicate that the belief changed in the same direction as the prior (so, 0.8 Iowa/ 0.2 New Mexico → 0.9 Iowa/ 0.1 New Mexico) and negative values indicate that the belief changed in the opposite direction as the prior (so, 0.8 Iowa/ 0.2 New Mexico → 0.7 Iowa/ 0.3 New Mexico). The magnitude of the evidence was equal to the absolute log likelihood ratio and the direction of the evidence was normalized so that positive values indicate that the evidence confirmed the prior and negative values indicate that the evidence disconfirmed the prior. For each block and each condition, a continuous heatmap was generated by summing Gaussian kernels with heights equal to the belief update, x coordinates equal to the prior belief certainty, and y coordinate equal to the evidence consistency for all observations. Heatmaps were subtracted from each other to account for baseline and sham stimulation.

C. Results

1. Sensations and side-effects of HD-tDCS

Participants most commonly reported that the stimulation felt like tingling, itching, pins and needles, burning, stinging, prickling, pressure, or poking (Figure 7). Fewer sensations were reported after the task was finished but before stimulation ended. Sensations intensities ranged from very mild to moderate. An "intense" rating was only used once to describe intense itching. Sensations were more intense 30 seconds into stimulation than several minutes into stimulation when participants finished the task (Figure 8). Sham was associated with milder sensations compared to RH-bias stimulation pre- task and compared to both LH-bias and RH-bias stimulation post-task, but these differences were not significant after correcting for multiple comparisons (Mann-Whitney U test: pre:RH>SH statistic=217.0, p=0.02, post:RH>SH statistic=155.5, p=0.02, and post:LH>SH statistic=178.5, p=0.03).

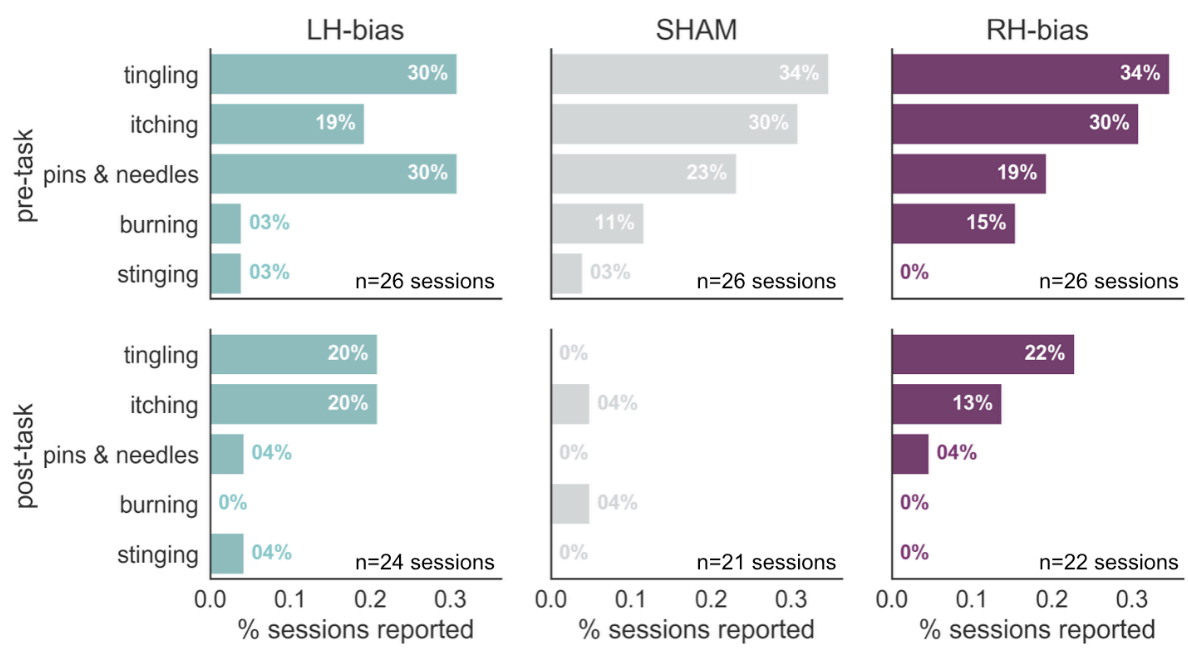

Figure 7: Reported sensations during stimulation. The pre-task sensations were reported approximately 30 seconds into stimulation and the post-task sensations were reported after

participants finished the task (if they finished before stimulation ended). Participants could report more than one type of sensation.

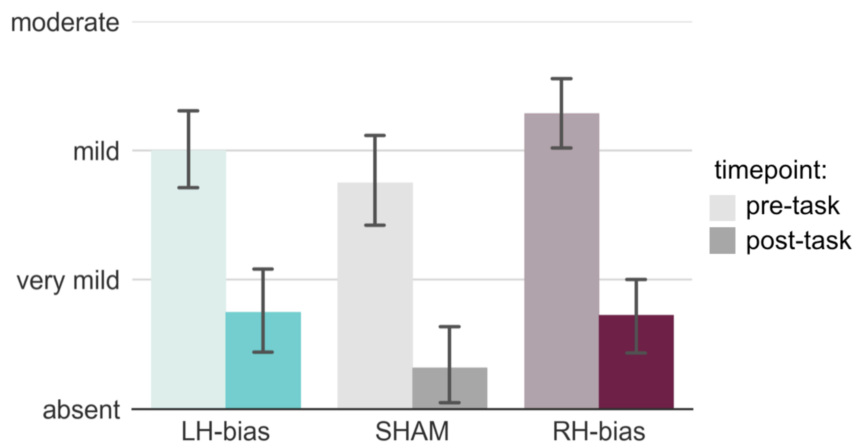

Figure 8: Sensation intensities before and after the state guessing task in each stimulation condition. Error bars represent 95% confidence intervals.

2. Demographic estimates

Every participant estimated the demographic makeup of the five U.S. states used in the experiment. Their estimates were used to calculate individualized ideal Bayesian posteriors (and, consequently, ideal slider positions and certainties) for every observation and every trial. The posteriors calculated with participants' demographic estimates were strongly correlated to those calculated with the true demographics from the 2010 census ($r=0.596$, $p<0.001$), but the relationship was certainty not perfect (Figure 9).

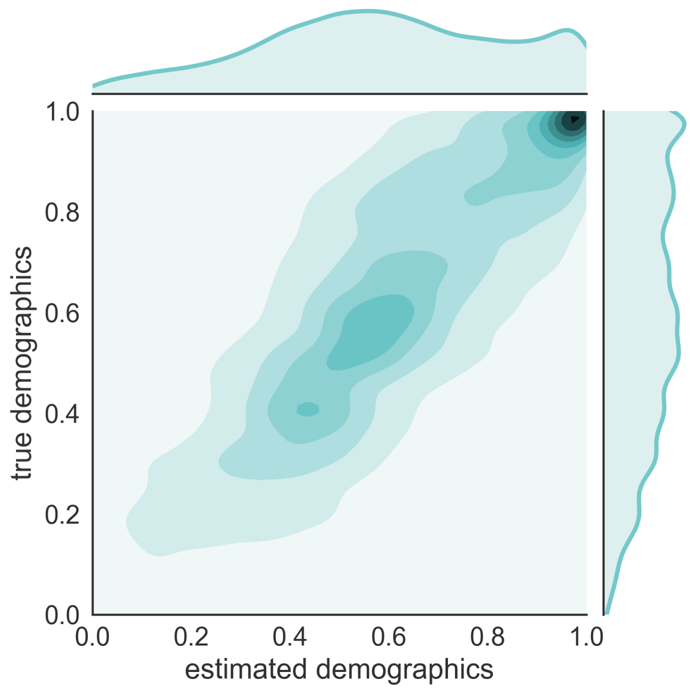

Figure 9: Comparison between the Bayesian posteriors calculated with participants' demographic estimates vs. the true 2010 demographics. Posteriors for observations at the middle and end of each sequence are predominately p=1, so only posteriors for the first 6 observations are included in this plot to aid visualization.

3. Amount of evidence collected

We predicted that the LH-bias stimulation would drive participants to become more certain of their guesses early on and thus stop evidence collection earlier than in the sham or RH-bias stimulation conditions. Contrary to our predictions, we found no evidence that the stimulation affected the amount of evidence participants collected compared to baseline (LH>baseline: -0.18 ± 1.15, SHAM>baseline: -0.03 ± 1.22, RH>baseline: -0.11 ± 1.31; LH>RH Wilcoxon signed-rank test: T=168.5, p=0.858).

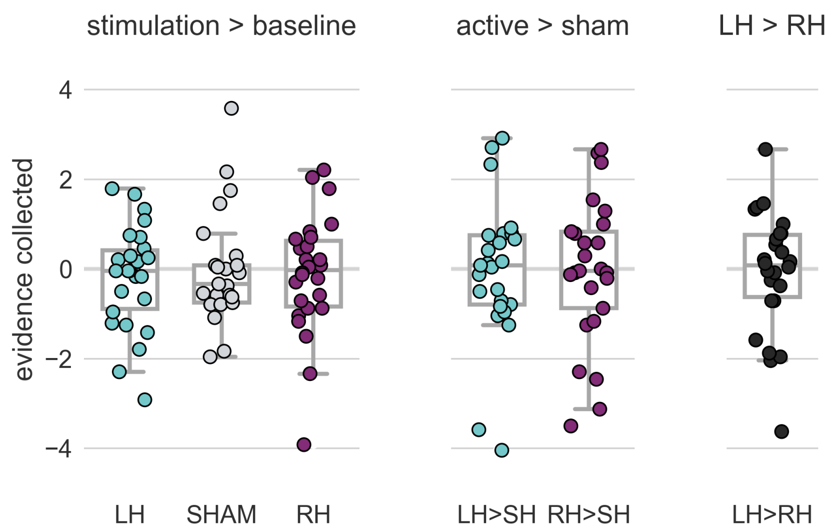

Figure 10: Amount of additional evidence collected compared to baseline and sham stimulation (n=26 participants).

We used multiple linear regression to examine the relationship between the amount of evidence collected and the stimulation conditions after taking additional factors into account, including individual differences in cortical current density (Figure 11). We predicted that RH-bias stimulation and LH-bias stimulation would be associated with more and less evidence collection compared to sham, respectively, and that these effects would be more pronounced with greater cortical current density. We found that greater cortical current density was associated with collecting more evidence in both the LH-bias and RH-bias conditions. The relationships between each participant's estimated cortical current density and average amount of evidence collected in each condition are presented in Figure 12. Consistent with the regression results, it is apparent that the two active stimulation sessions are associated with slightly less evidence collection on average (represented by the dotted lines), but evidence collection increases as cortical current density increases.

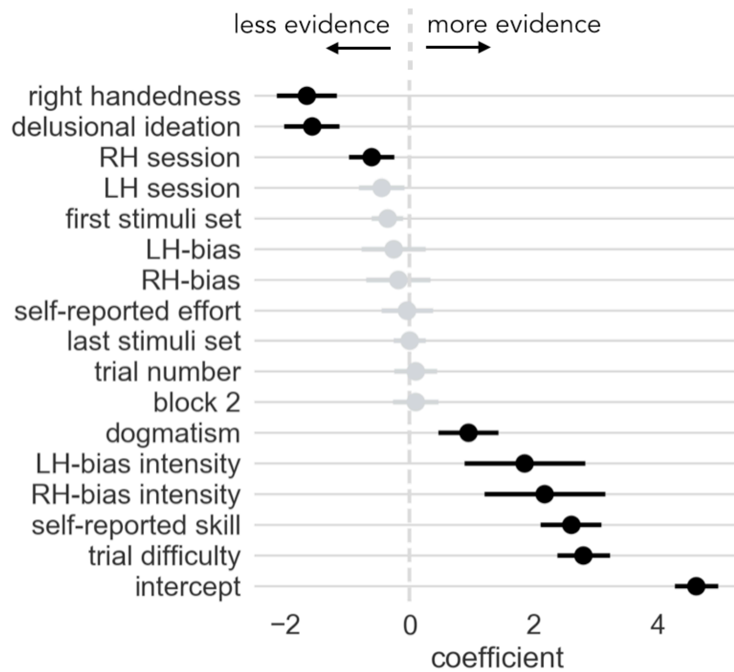

Figure 11: Linear regression results for amount of evidence collected. Values in black are significant after correcting for 17 comparisons with a Bonferroni correction ($t>2.98$, $p<0.003$).

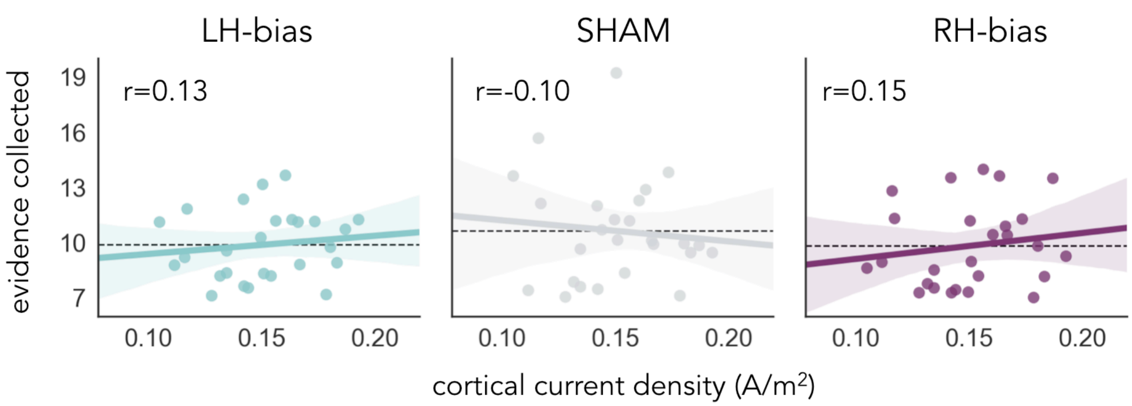

Figure 12: Relationships between cortical current density and the amount of evidence collected in each stimulation condition. Each point represents one participant and only the data from the second block (during stimulation) is presented. Shaded regions denote 95% confidence intervals. Note: cortical current density is not expected to be correlated with any behavioral measure in the sham condition, since no stimulation is delivered.

Several other factors influenced evidence collection. As expected, participants collected more evidence in trials that were more difficult (when the evidence did not point strongly toward one state or the other). Several participant measures also accounted for differences in evidence collection. Higher dogmatism scores and higher self-reports of skill at the task where associated with collecting more evidence, while more extreme right-handedness and greater delusional ideation scores were associated with collecting less evidence.

4. Threshold for stopping evidence collection

We predicted that, compared to baseline and sham stimulation, LH-bias stimulation would be associated with a requiring a lower threshold of evidence and RH-bias stimulation would be associated with requiring a higher threshold of evidence before stopping evidence collection. To test this prediction, we computed the ideal Bayesian posterior for the state that the participant selected at the final evidence presentation. We restricted the analysis to trials in which participants decided to stop collecting evidence (that is, the trials in which the participant ran out of evidence were discarded) and trials in which the final decision was aligned with the preponderance of evidence (that is, the trials in which the participants selected the state with an estimated posterior < 0.5 were discarded). There were no differences between evidence thresholds compared to baseline (LH>baseline: -0.008 ± 0.058, SHAM>baseline: -0.001 ± 0.049, RH>baseline: 0.002 ± 0.042; LH>RH Wilcoxon signed-rank test: T=174.0, p=0.969).

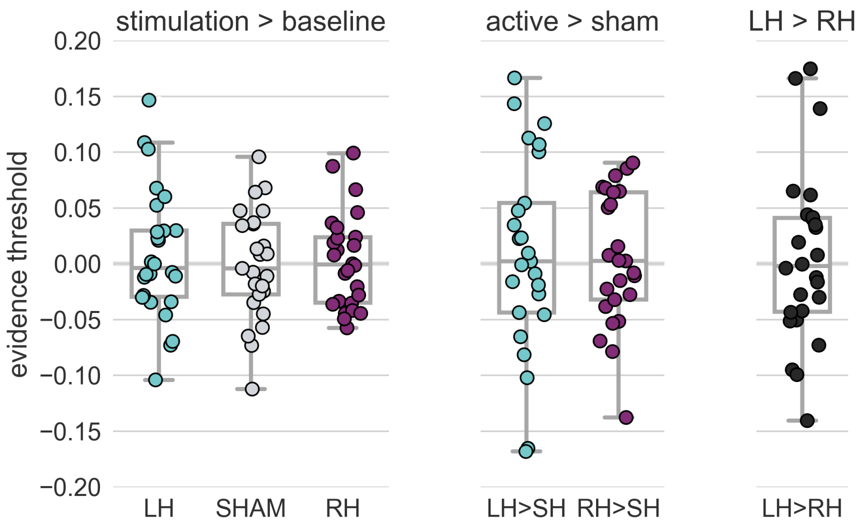

Figure 13: Threshold for stopping evidence collection compared to baseline and sham (n=26 participants).

We ran a regression analysis with the same design matrix to examine the influence of additional factors on participants' evidence thresholds. We found that, as expected, greater RH-bias intensities were associated with higher thresholds to stop evidence collection (Figure 14). The positive relationship between stimulation intensity and evidence threshold in the RH-bias condition is illustrated in Figure 15. We also found that, similar to before, trial difficulty and right-handedness were associated with lower evidence thresholds and self-reported skill was associated with higher thresholds.

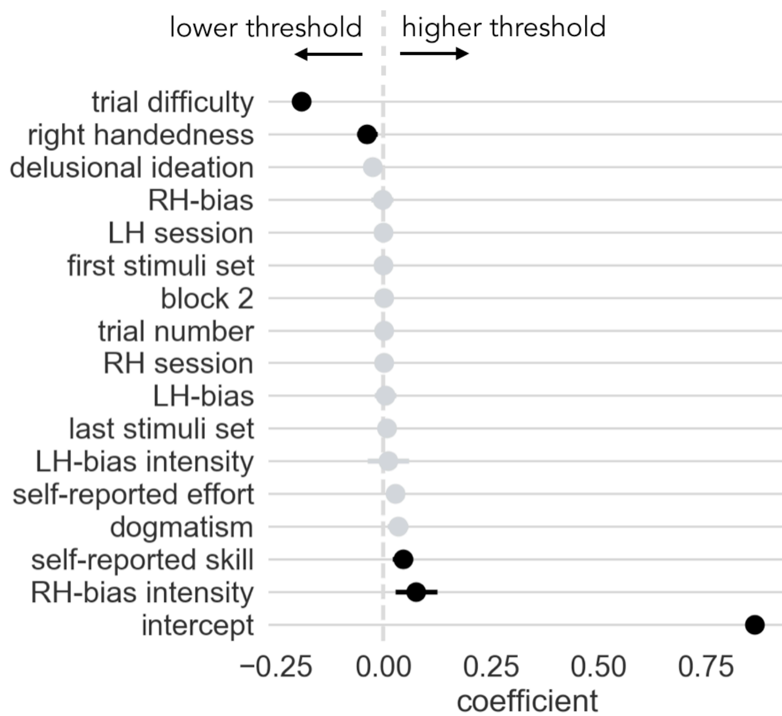

Figure 14: Linear regression results for threshold used to stop evidence collection. Values in black are significant after correcting for 17 comparisons with a Bonferroni correction ($t>2.98$, $p<0.003$).

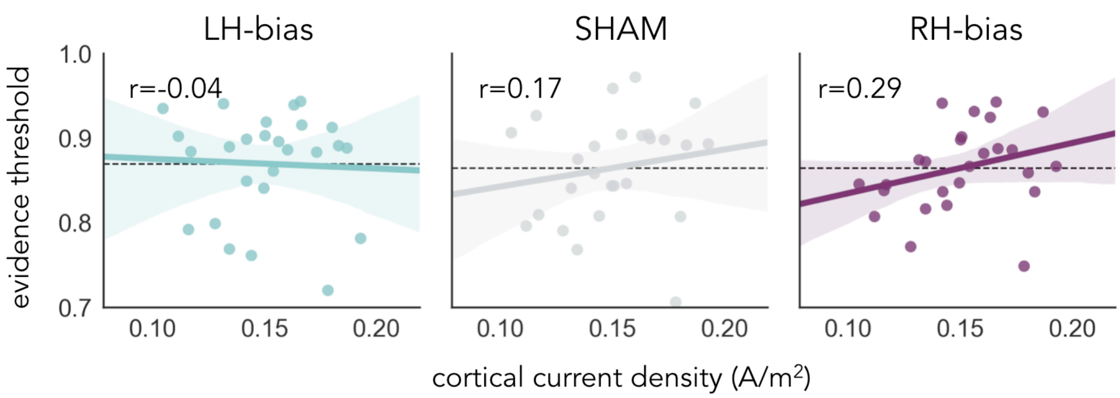

Figure 15: Relationships between cortical current density and evidence thresholds in each stimulation condition. Each point represents one participant and only the data from the second block (during stimulation) is presented. Shaded regions denote 95% confidence intervals. Note: cortical current density is not expected to be correlated with any behavioral measure in the sham condition, since no stimulation is delivered.

5. Certainty throughout the evidence presentation

Certainty was operationalized as the distance away from the midpoint of the digital slider. Certainty values ranged from 0 at the slider midpoint and increased linearly up to 0.5 at either end of the slider. We predicted that participants would endorse more certain state guesses under LH-bias stimulation compared to RH-bias stimulation, especially early on in each trial before much evidence was accumulated. Only the certainties associated with the first 7 residents of each trial were analyzed since all participants observed those residents. The expected certainty -- equal to the certainty associated with the ideal Bayesian posterior -- was computed for each observation and subtracted from participants' certainty values in order to account for differences in evidence strength across trials. After normalizing for expected certainty, baseline, and sham stimulation, the certainty trajectories under LH-bias and RH-bias stimulation were compared. Compared to RH-bias stimulation, LH-bias stimulation was associated with greater certainty, but only early on in the trial (Figure 16 A). A similar pattern emerged when the data was grouped by stimuli ID rather than participant ID (Figure 16 B). However, none of the certainty differences between LH-bias and RH-bias stimulation were significantly different from zero after using a Bonferonni correction to correct for multiple comparisons.

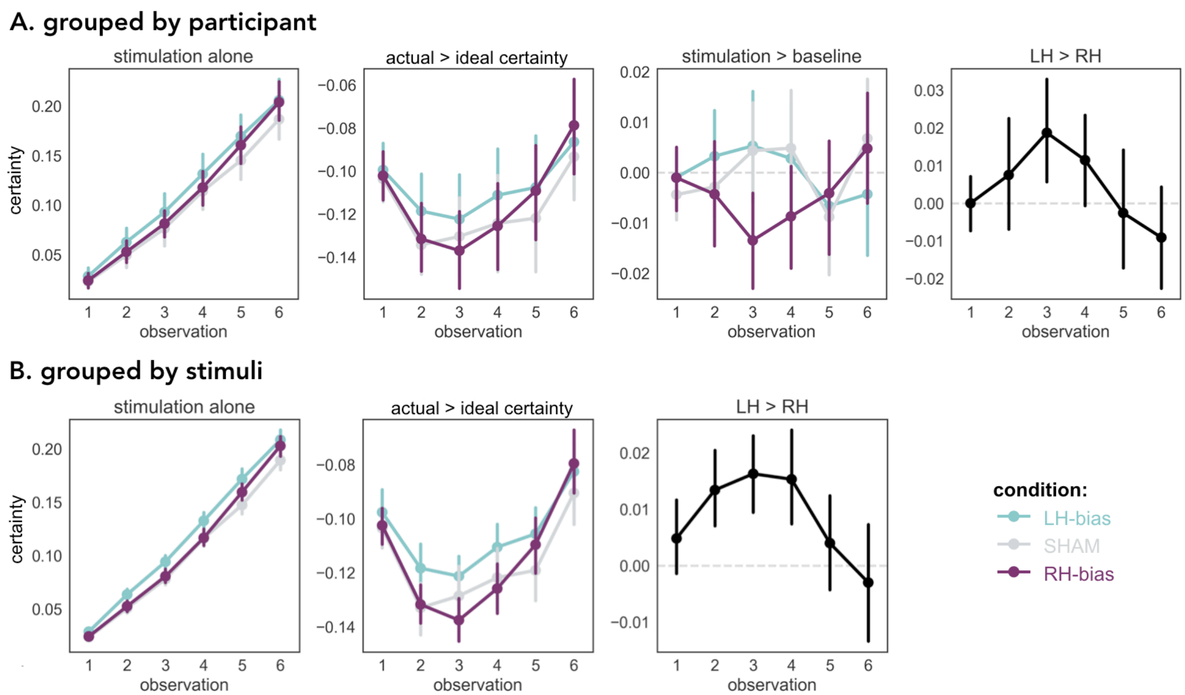

Figure 16: Certainty associated with the first 6 observations, normalized by expected certainty, baseline, and sham. A. Each point represents the mean across participant means (n=26). B. Each point represents the mean across stimuli means (n=72 sequences). Error bars represent standard error.

We used trajectory regression analysis to determine how, if at all, tDCS affected how certainty changed throughout the course of the trial. First, the ideal certainty (defined as the certainty associated with the ideal Bayesian posterior) for each observation was subtracted from the certainty values in order to normalize certainty trajectories across trials. The normalized certainty trajectories were clustered into two groups using k-means clustering (Figure 17). The trials in the first cluster, named "ideal", were associated with trajectories that were close to the ideal certainty levels. On average, the certainties were slightly lower than ideal at the very beginning of the trial and slightly higher than ideal later on. The trials in the second group, named "uncertain", were associated with lower certainty than ideal throughout the first seven observations.

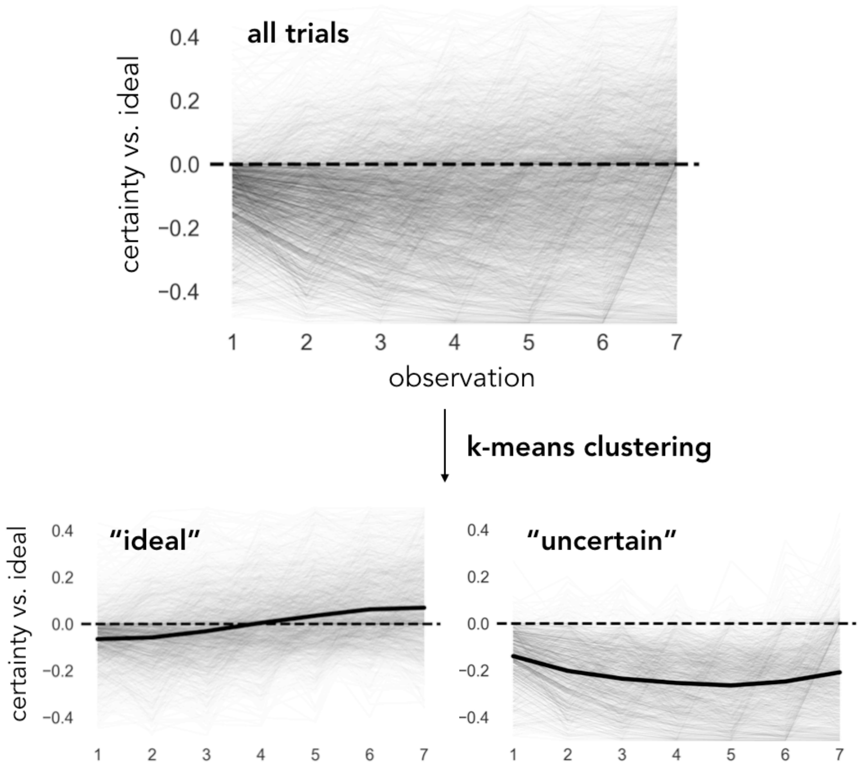

Figure 17: K-means clustering of certainty trajectories in preparation for trajectory analysis.

Multinomial logistic regression was used to determine what factors made it more likely that any given trial would be classified as an "uncertain" trajectory as opposed to "ideal" trajectory (Figure 18). We found that trial difficulty was more predictive of "ideal" trajectories than "uncertain" trajectories. Only two participant measures were significantly associated with trajectory type: self-reported ratings of effort on the task and higher endorsement of delusional ideas were both more predictive of the "uncertain" trajectories than "certain" trajectories. Finally, we found that greater RH-bias intensity (defined as greater cortical current density in the RH-bias stimulation condition) was more predictive of the "uncertain" trajectory than the "ideal" trajectory. To visualize the relationship between trajectory type and cortical current density, Figure 19 shows scatter plots of the fraction of "uncertain" trajectory types for each participant compared to their estimated cortical current density for each stimulation condition.

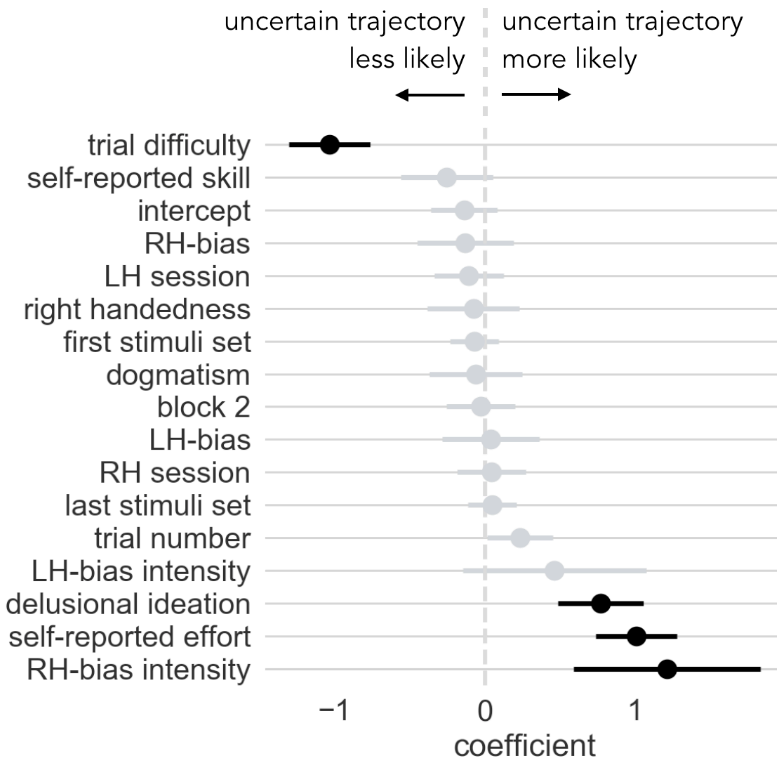

Figure 18: Linear regression results for threshold used to stop evidence collection. Values in black are significant after correcting for 17 comparisons with a Bonferroni correction (t>2.98, p<0.003).

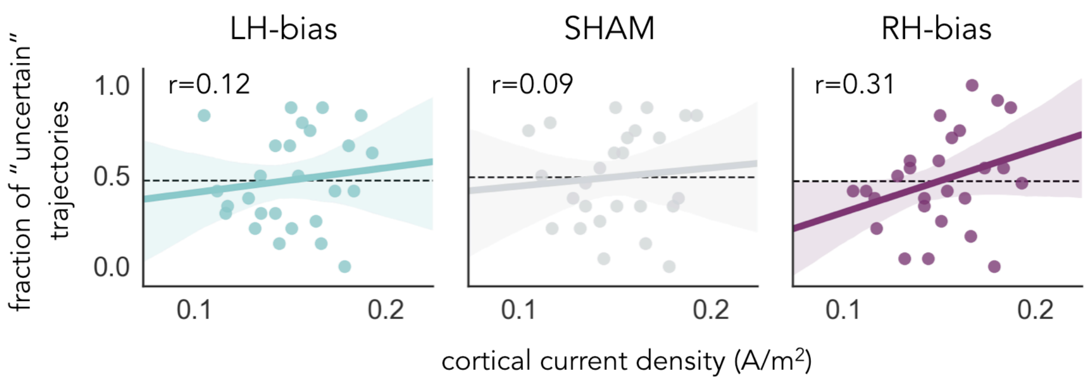

Figure 19: Relationships between cortical current density and fraction of "uncertain" trajectory types. Each point represents one participant and only the data from the second block (during stimulation) is presented. Shaded regions denote 95% confidence intervals.

6. **Belief updating after belief-evidence conflicts**

Based on the hemispheric lateralization framework, we expected that tDCS would have the most divergent effects in situations when evidence contradicted strong prior beliefs. In these situations, we predicted that LH-bias stimulation would bias participants toward maintaining a high level of certainty so participants would be less likely to change their beliefs after receiving conflicting evidence. Conversely, we predicted that RH-bias stimulation would make participants more responsive to conflicting evidence and more likely to update their beliefs until they were more in line with the new evidence. To test these predictions, we selected time points in which 1) participants had a strong prior belief that a particular state was selected (p<0.1 or p>0.9) and 2) they observed a resident that was 3x or more likely to come from the opposite state. Participants' belief changes in these instances were compared between blocks and between conditions. Belief change was operationalized as the distance between the slider positions before vs. after receiving the conflicting evidence and was normalized to participants' prior belief such that positive belief changes indicate the participant became more certain of their state guess and negative changes indicate the participant became less certain (or moved closer to the opposite state). In all blocks, conflicting evidence was associated with moving the slider closer toward the opposite state, in the same direction as the conflicting evidence (Figure 20). There were no differences in belief updates in the sham or RH-bias condition compared to baseline, but LH-bias was associated with a significantly larger average belief update compared to baseline (t=3.03, p=0.003).

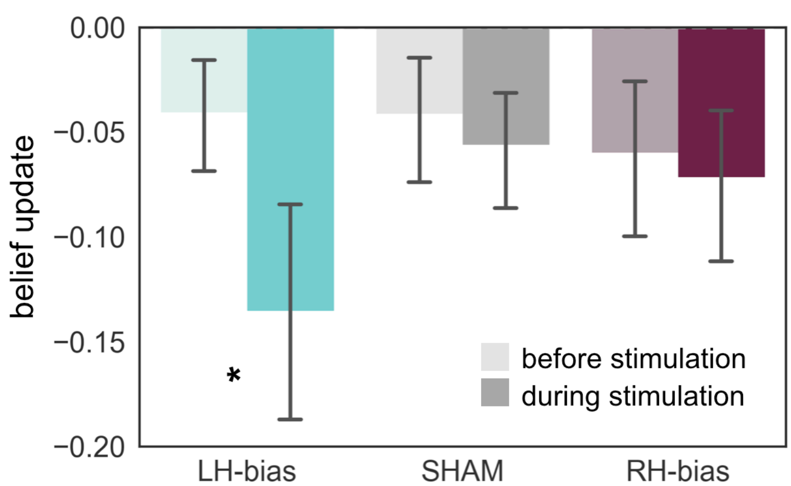

Figure 20: Change in belief after receiving conflicting evidence. Each bar represents the average change in the slider position for observations that conflicted a strongly held belief. The number of observations per bar ranges from n=97 to n=158. Negative belief change values indicate that the slider was moved in the opposite direction from the prior (that is, if a participant initially thought state A was selected, but then moved the slider more toward state B). Error bars represent standard error.

One important caveat of this analysis is that it is sensitive to the cutoffs used for prior belief strength and evidence conflict. To obviate the need for cutoffs, participants' belief updates were computed over the full range of prior belief certainties (from completely uncertain to completely certain) and evidence consistencies (from highly disconfirmatory to highly confirmatory of the prior). For each block and each condition, a continuous heatmap was generated by summing Gaussian kernels with heights equal to the belief update, x coordinates equal to the prior belief certainty, and y coordinates equal to the evidence consistency for all observations. According to our framework, the left hemisphere is biased toward reducing and maintaining certainty, so we predicted that LH-bias stimulation would make participants update their beliefs more when certainty is low but less when certainty is high compared to baseline (Figure 21). Similarly, we proposed the right hemisphere applies a "cognitive brake" when evidence conflicts strong beliefs, so we predicted that, compared to baseline, RH-bias stimulation would make participants more likely to backtrack on their beliefs in cases where they were certain and received disconfirmatory evidence.

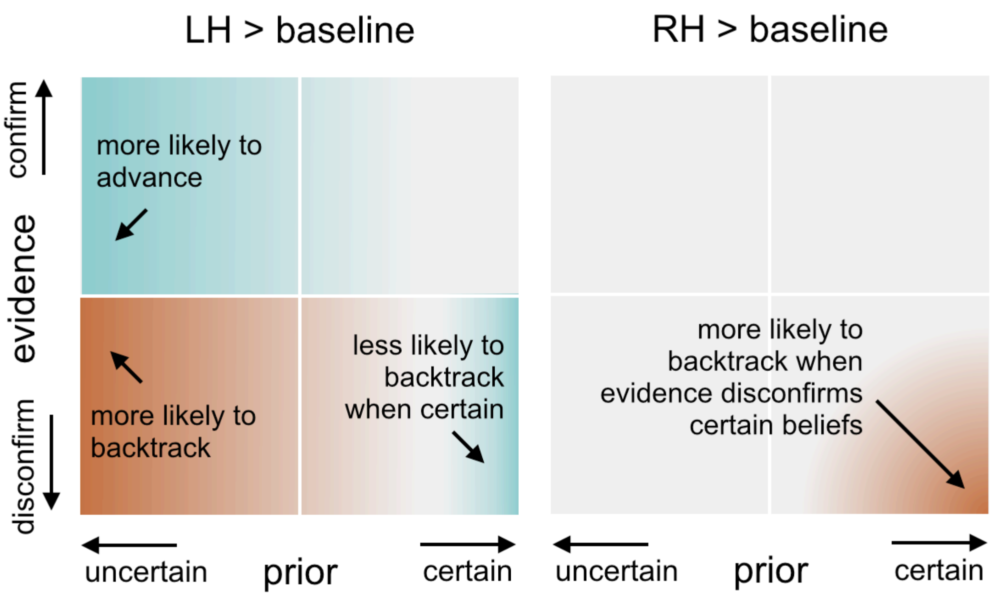

Figure 21: Predictions of the effect of LH-bias and RH-bias stimulation on belief updating according to prior belief certainty and evidence consistency.

Compared to baseline, LH-bias stimulation was generally associated with more extreme updates, regardless of prior belief certainty. We found no evidence that RH-bias stimulation was associated with larger backtracks when evidence conflicted strong prior beliefs. In fact, compared to LH-bias stimulation, RH-bias stimulation was associated with less backtracking after belief-evidence conflicts (Figure 22).

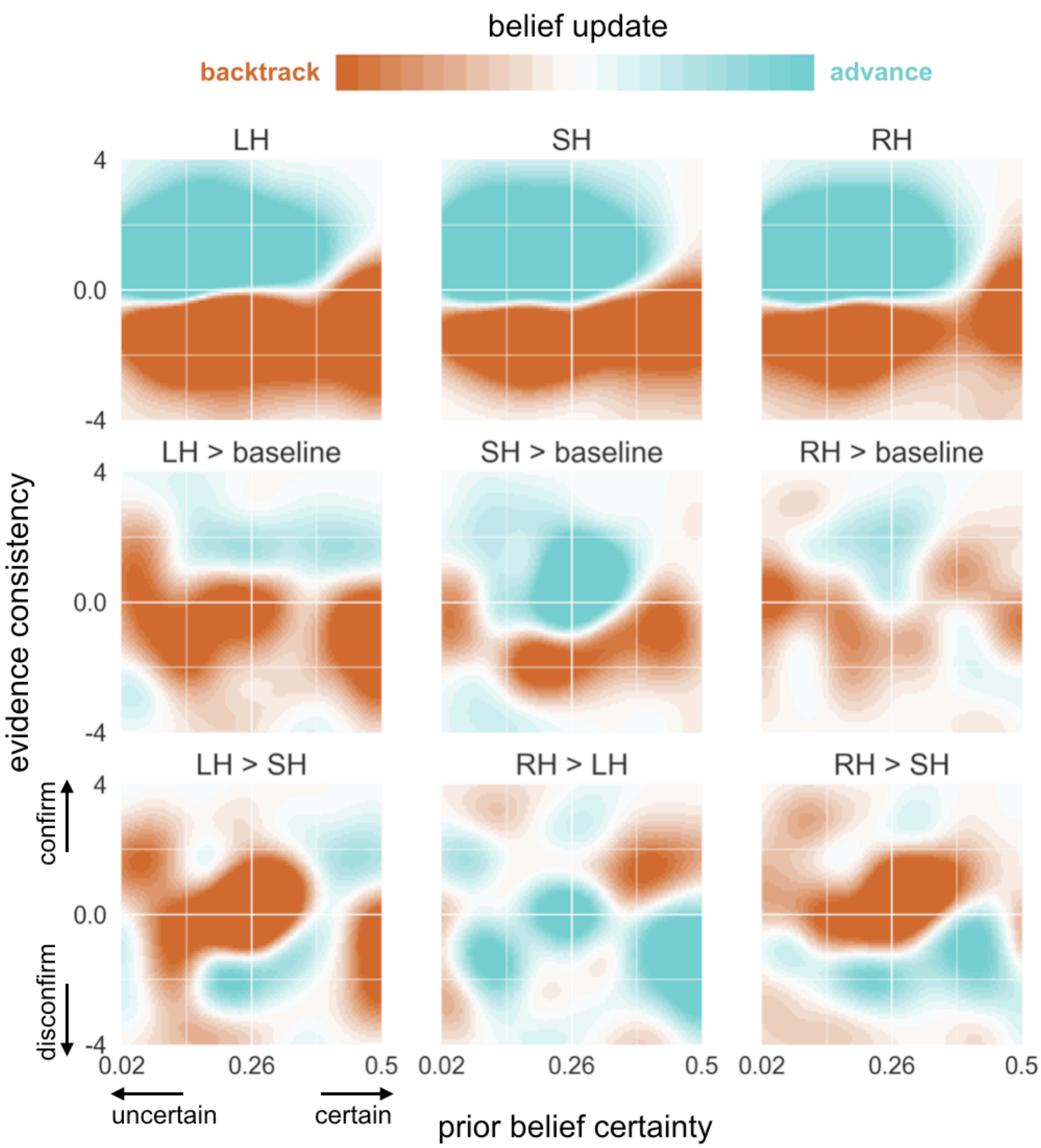

Figure 22: Normalized belief updates according to prior certainty and evidence consistency. Areas in blue denote regions where the slider was updated in the same direction as the prior and areas in red denote regions where the slider was updated in the opposite direction (that is, a backtrack).

Finally, the distribution of observations across belief certainties and evidence consistencies were compared between stimulation conditions (Figure 23). Compared to baseline, all three stimulation conditions were associated with fewer uncertain observations and more observations with moderate to high certainty. Likewise, and consistent with our predictions, LH-bias stimulation was associated with fewer uncertain observations and more moderately to highly certain observation compared to RH-bias stimulation.

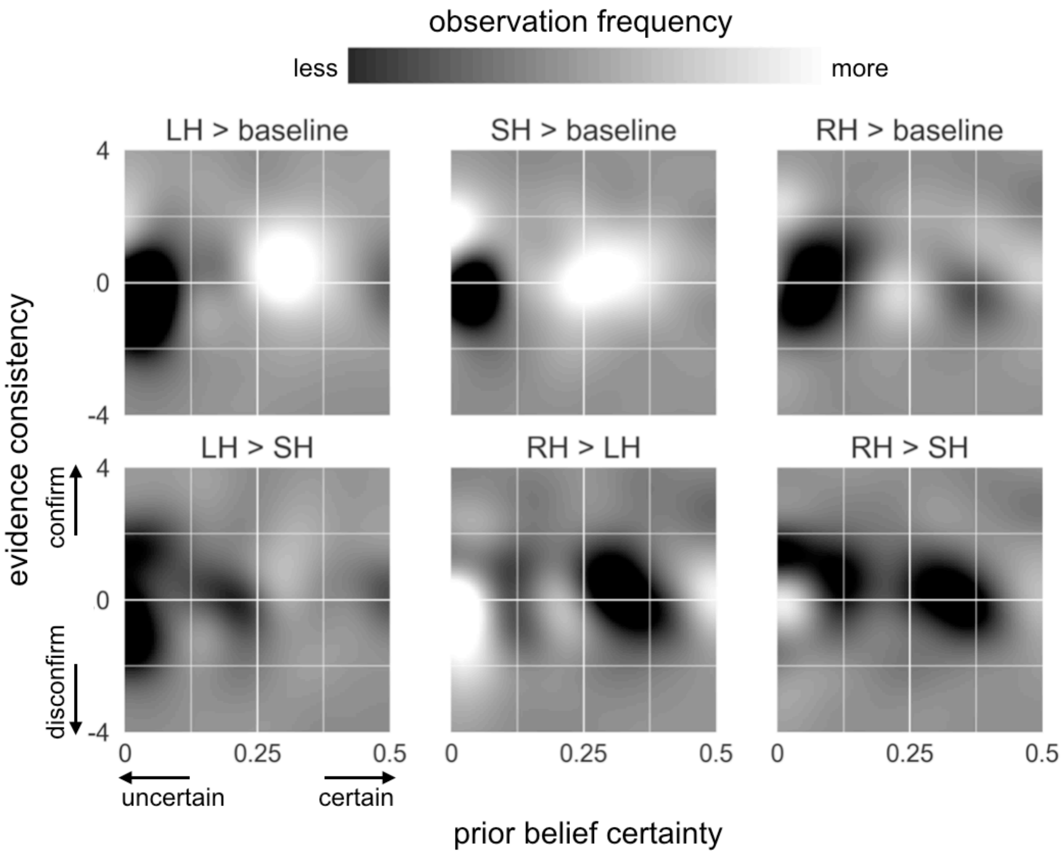

Figure 23: Distribution of observations according to prior certainty and evidence consistency. Areas in white and black denote regions or more or fewer observations, respectively.

D. Discussion

One of our primary predictions was that LH-bias stimulation would drive participants toward being more certain, which would result in participants collecting less evidence and adopting a lower threshold to stop evidence collection. Conversely, we predicted that, under RH-bias stimulation, participants would exhibit less certainty and would require higher evidentiary thresholds for stopping evidence collection. Our results were consistent with our predictions for RH-bias stimulation: greater cortical current density in the RH-bias condition was significantly associated with 1) collecting more evidence, 2) adopting a higher threshold for stopping evidence collection, and 3) less certainty than ideal throughout the beginning of the trial. Although there was some indication that LH-bias stimulation was associated with greater certainty (Figure 16 and Figure 23), we also found that greater LH-bias intensity was associated with more evidence collection, which was contrary to our predictions. We did not find any significant effects of LH-bias intensity on evidence thresholds or certainty trajectories, although the directions of the effects were also opposite of our predictions.

One possibility for the discrepancy between our predictions and the results for the LH-bias stimulation is that requiring participants to view at least six residents before stopping evidence collection concealed early differences between conditions and perhaps differences would have emerged if participants were allowed to stop evidence collection at any time. As shown in Figure 16, the largest differences in certainty between LH-bias stimulation and RH-bias stimulation occurred early on in the trial when participants were not allowed to stop collecting evidence. By the time participants were allowed to stop collecting evidence after observation 6, the certainties for the two stimulation conditions were similar. It's possible that LH-bias stimulation would be associated with less evidence collection if participants were given the option to stop evidence collection early on in the trial when differences in certainty were more pronounced.

The reason participants were required to observe at least six residents was to ensure that a full dataset could be analyzed for at least a subset of observations. Allowing participants to stop evidence collection at any time would hinder direct comparisons of observations across trials, participants, blocks, and conditions, partly because it would produce unequal sample sizes, but also because it could produce a dropout bias. For example, the effect of stimulation on participants' certainty trajectories could not be analyzed because there would be a drop out of trials with each additional observation and the drop outs would be confounded with certainty. Six residents were chosen as the minimum number of residents based on initial pilot data in which participants collected more residents on average. Participants may have collected fewer observations in this experiment compared to the pilot experiment due to different task demands, differences in trial difficulty, or greater familiarization with the task. To mitigate floor effects in the future, the number of required observations could be reduced to only 2, 3, or 4 residents, the trial difficulty could be increased to encourage participants to collect more evidence, bonuses based on trial accuracy could be given to motivate participants to be more certain before making a final guess, or an observation minimum could be applied to only half of the trials.

Secondly, we predicted that RH-bias stimulation would make participants more sensitive to conflicts between evidence and beliefs and more likely to revise strong beliefs after receiving conflicting evidence. We found no evidence that RH-bias stimulation made participants more sensitive to belief-evidence conflict and, in fact, we found that participants were less likely to revise their strongly held beliefs after receiving conflicting evidence compared to LH-bias stimulation. LH-bias stimulation was associated with significantly larger backtracks in strongly held beliefs after receiving conflicting evidence compared to baseline, while there were no significant differences in backtracks in RH-bias stimulation compared to baseline. Likewise, comparing belief updates over the full range of prior certainties and evidence types showed that LH-bias stimulation was associated with larger backtracks compared to RH-bias stimulation when prior beliefs were strong and evidence was conflicting.

It is possible that RH-bias stimulation did not produce greater belief updating after belief-logic conflicts because this task did not engender the same sort of belief-logic conflicts as those reported in the literature. Although we use the term “strong belief” to denote slider positions near either end of the track, these are not strong beliefs in the normal sense of the phrase. Participants’ “certain beliefs” in this experiment pale in comparison to their “certain beliefs” in real life, especially compared to beliefs about politics or religion, for example. In short, it is possible that the belief-evidence conflicts participants encountered in this experiment did not produce enough conflict to trigger a right-lateralized “cognitive brake.” To test this possibility, stimuli with more meaning and real-world applicability were developed and used in a second tDCS experiment.

Finally, we found that participants who scored higher on the delusional ideation scale collected less evidence on average, which is consistent with the patient literature that finds that delusional patients exhibit a strong “jump to conclusions bias” and collect fewer beads than controls in the beads task. Interestingly, greater delusional ideation was also associated with the “uncertain” certainty trajectory. It’s unclear why delusional ideation is associated with both less evidence collection and less certainty throughout the evidence presentation (while having no effect on evidence threshold), but it is possible that participants with higher delusional ideation have a poorer meta-cognitive assessment of their certainty levels or, perhaps even more likely, do not faithfully indicate their true certainty levels by moving the slider.

III. Experiment 2: Effects of HD-tDCS on real-world reasoning

A. Rationale

In Marinsek et al. (2014), we argued that subtle lateralized processing – if present – may contribute toward a more flexible reasoning system that can easily adapt to contextual reasoning demands. For example, situations requiring caution and a high standard of evidence may benefit from biasing right PFC networks while situations that favor liberal inference making and penalize uncertainty may benefit from biasing left PFC networks. If this is the case, it follows that if we artificially bias activity to the left PFC or right PFC, we should observe changes in how participants reason in realistic, nuanced scenarios.

One drawback of the state guessing task is that it may be too abstract, too inconsequential, or too far removed from real-world reasoning to elicit the reasoning processes that we make predictions about. Although we took some steps to make the task more ecologically valid than the original beads task (for example, by expanding the number of evidence types and requiring participants to rely on their background knowledge to make inferences), the task may fall short of capturing real-world, flexible reasoning. In particular, the task may fail to produce salient belief-evidence conflicts that we predicted would lead to divergent behavior under LH-bias and RH-bias stimulation. Since the inferences and guesses that participants formed were likely much more pliable and inconsequential than participants' real beliefs, it's possible that disconfirmatory evidence would produce less conflict in the state guessing task than in the real world.

In this experiment, we again asked participants to make decisions based on their background knowledge and new, sometimes disconfirmatory, evidence, but used reasoning scenarios that are more deeply embedded in real-world contexts. The first set of scenarios required participants to judge whether criminal defendants were guilty of a crime based on evidence from real criminal investigations and the second set of scenarios required participants to vote on hypothetical ballot measures after reading arguments in favor or in opposition to them. Both of these scenarios have important consequences in the real world. Voters must weigh conflicting evidence when deciding whom to elect to powerful positions in government and whether or not to vote propositions into law. Likewise, jurors are required to make decisions based on a great deal of conflicting evidence and risk convicting an innocent person or letting a guilty person go free. Even though people do not vote or sit on juries in their daily lives, many people reason about similar scenarios, albeit to a lesser extent, on a daily basis when forming opinions and judgments about personal or current events.

In a final task, participants had to decide whether a news headline described a real event that actually happened or a fake event that was made-up. To do so, participants had to judge how likely an event was true based on their knowledge of the world (that is, they were essentially tasked with estimating the likelihood ratio in Bayes' rule). Although these judgments may appear to be easy, they can be quite difficult since real, verifiable events may be bizarre and made-up events may appear plausible, especially if they are consistent with one's worldview. Indeed, the prevalence of erroneous beliefs due to accepting fake news as real is widespread: a 2017 Public Policy Poll reported that 32% of Americans (including 60% of Trump supporters) believe that millions of people voted illegally for Hillary Clinton in the 2016 presidential election, 42% of Americans (including 73% of Trump supporters) believe George Soros paid people to protest Trump, and 14% of Trump supporters believe Hillary Clinton is involved in a child sex ring run out of a pizza restaurant in Washington D.C. All of these stories are linked to fake news articles that have been thoroughly debunked. The goal of the fake news task wasn't necessarily to test how well participants can discriminate between real and fake headlines, but instead to explore whether biasing neural activity to the left or right hemisphere biased participants' criterion for believing new information.

B. Methods

1. Participants

24 individuals (16 females, mean age: 21.4, age range: 19 to 37) participated in the transcranial direct current stimulation experiment. All participants who participated in the first tDCS experiment were invited to participate in this experiment, and 17 out of 26 participants chose to participate again. The remaining participants were recruited from a pool of individuals who had passed a screening process that ensured they could follow task instructions adequately and they could safely be scanned with MRI and stimulated with tDCS and TMS (the selection process is described in detail in section II.A.1). Participants were paid $20 per hour for their participation in the experiment.

2. Tasks

Participants completed three tasks during stimulation, each of which are described below.

Criminal Court Case Scenarios: Participants were instructed to imagine that they were sitting on a jury and they needed to decide whether a defendant was guilty or not guilty. At the beginning of each trial, participants were given a very concise overview of the criminal court case, for example: "George Allen stands trial for raping and murdering a young court reporter in her home" (Figure 24). Participants then used a digital slider to report their opinion to the question, "Is George Allen guilty of rape and murder?". The leftmost point of the slider was labeled "definitely not guilty" and the rightmost point was labeled "definitely guilty." The mouse cursor and slider were positioned in the center of the slider track at the beginning of every trial. Participants used the mouse to move the slider back and forth along the track and clicked the mouse button to submit their response. Participants were then presented with crime scene details. Three pieces of evidence suggested that the defendant was innocent (for example, "A fingerprint thought to belong to the attacker did not match any of Allen's fingerprints.") and three pieces of evidence suggested that the defendant was guilty (for example, "Semen found at the crime scene contained antigens that are similar to the antigens in Allen's blood.") The order of the crime scene evidence was counterbalanced across trials and across participants. After each piece of evidence was presented, participants again reported their opinion of the defendant's guilt with the digital slider. After all six pieces of evidence were presented, participants were given a two-alternative forced choice between deciding whether the defendant was "guilty" or "not guilty." Participants also rated how confident they were that they made the correct decision on a scale from 1-5, where: 1=low, 2=somewhat low, 3=moderate, 4=somewhat high, and 5=high.

Figure 24: Example of stimuli from criminal court case scenarios.

Crime details were taken from the cases published by the Innocence Project (www.innocenceproject.org). In each case, the defendant was convicted of the crime but the conviction was later overturned when DNA evidence proved that the defendant was innocent. These cases made for excellent stimuli because the crime scene evidence was often conflicting and ambiguous. 48 court case scenarios were created with 6 pieces of evidence each. A separate group of 48 participants completed the court case task on testing computers without receiving any stimulation. The ratings of these participants were used to determine which pieces of evidence were the most convincing by solving for the likelihood ratio in Bayes' rule:

$$\frac{P(evidence|guilty)}{P(evidence|innocent)} = \frac{P(guilty|evidence)}{P(innocent|evidence)} \div \frac{P(guilty)}{P(innocent)}$$

Equation 2: Adaptation of Bayes' rule to estimate persuasiveness of crime evidence.

Like before, the slider position before the evidence was presented was used to estimate the prior probabilities of P(guilty) and P(innocent), and the slider position after the evidence was presented was used to estimate the posterior probabilities. The natural log of the likelihood ratio was used as an estimate of how convincing a crime scene detail was. Positive log likelihood ratios indicate that the evidence made participants more likely to give a guilty verdict, and negative log likelihood ratios made participants more likely to give a non-guilty verdict.

The 48 criminal court cases were divided into 3 groups with 16 court cases each. There was very little variation between participants' initial beliefs, so the court cases were sorted into groups based on how the evidence as a whole was convincing of guilt or innocence (as measured by the average log likelihood ratio associated with each court case). The three stimuli groups did not significantly differ in prior opinion, final opinion, overall log likelihood ratio, proportion of guilty verdicts, confidence, or reaction time.

Voter Ballot Scenarios: Participants were instructed to imagine that they were at the ballot box and they needed to vote whether or not to pass a ballot measure and make it a law. In each trial, participants were given a brief overview of a proposed law and asked to indicate their current opinion with a digital slider (Figure 25). For example, participants were told “You are voting on a law that will lower the drinking age to 18.” and used the digital slider to report their opinion to the question, “Should the drinking age be lowered to 18?”. In every trial, the leftmost point of the slider was labeled “definitely no” and the rightmost point was labeled “definitely yes.” The mouse cursor and slider were positioned in the center of the slider track at the beginning of every trial. Participants used the mouse to move the slider back and forth along the track and clicked the mouse button to submit their response. After indicating their baseline opinion, participants were asked to rate how much they had thought about the issue before that day on a 1-6 scale, where: 1=not at all, 2=very little, 3=a little bit, 4=a fair amount, 5= a good amount, and 6=a whole lot. Next, participants were shown two arguments in favor of the proposed law and two arguments in opposition to it. The order of the arguments was counterbalanced across trials and across participants. Participants reported their opinion of the ballot measure with the digital slider after each argument was presented. After all four arguments were presented, participants were given a two-alternative forced choice between voting “yes” or “no” on the ballot measure. Participants also rated how confident they were that they made the best decision on a scale from 1-5, where: 1=low, 2=somewhat low, 3=moderate, 4=somewhat high, and 5=high.

Figure 25: Example stimuli from ballot measure task.

Ballot measures and associated arguments were gathered from procon.org, ballotpedia.org, and voter information guides. 48 ballot measures were created with 6 arguments each (3 pro arguments and 3 con arguments). A separate group of 48 participants completed the voter ballot task on testing computers without receiving any stimulation. The ratings of these participants were used to determine which arguments were most convincing by solving for the likelihood ratio in Bayes' rule:

$$\frac{P(argument|yes)}{P(argument|no)} = \frac{P(yes|argument)}{P(no|argument)} \div \frac{P(yes)}{P(no)}$$

Equation 3: Adaptation of Bayes' rule to estimate the persuasiveness of each argument.

The slider position before an argument was presented was used to estimate the prior probabilities of P(yes) and P(no), and the slider position after an argument was presented was used to estimate the posterior probabilities, P(yes|argument) and P(no|argument). The natural log of the likelihood ratio was used as an estimate of how persuasive an argument was. Positive log likelihood ratios indicate that the argument made participants more likely to vote yes on the ballot measure, and negative log likelihood ratios made participants more likely to vote no on the ballot measure. To reduce the time of the task during stimulation, the least convincing pro argument (with the lowest average log likelihood ratio) and least convincing con argument (with the highest average log likelihood ratio) were excluded for each ballot measure, resulting in four arguments per ballot measure.

Ballot measures varied greatly based on 1) how much people knew about the topic, cared about the topic, and considered the topic beforehand (for example, participants knew more about abortion than returning to the gold standard), 2) participants' initial opinions (for example, most participants were strongly against allowing performance enhancing drugs in sports and strongly for allowing gay marriage), and 3) how the arguments changed participants' opinions (for example, participants were more likely to vote for cell phones being banned in schools and less likely to vote for using flavored milk in school lunches after reading all arguments). An iterative agglomerative clustering method was developed to cluster the ballot measures into 16 clusters with 3 scenarios each based on the three dimensions described above (Figure 26). For each cluster, the three ballot measures were randomly assigned to one of three stimuli groups. The three stimuli groups did not significantly differ in prior opinion, final opinion, overall log likelihood ratio, proportion of yes votes, confidence, reaction time, care rating, knowledge rating, or consideration rating. The three stimuli groups were assigned to three stimulation sessions.

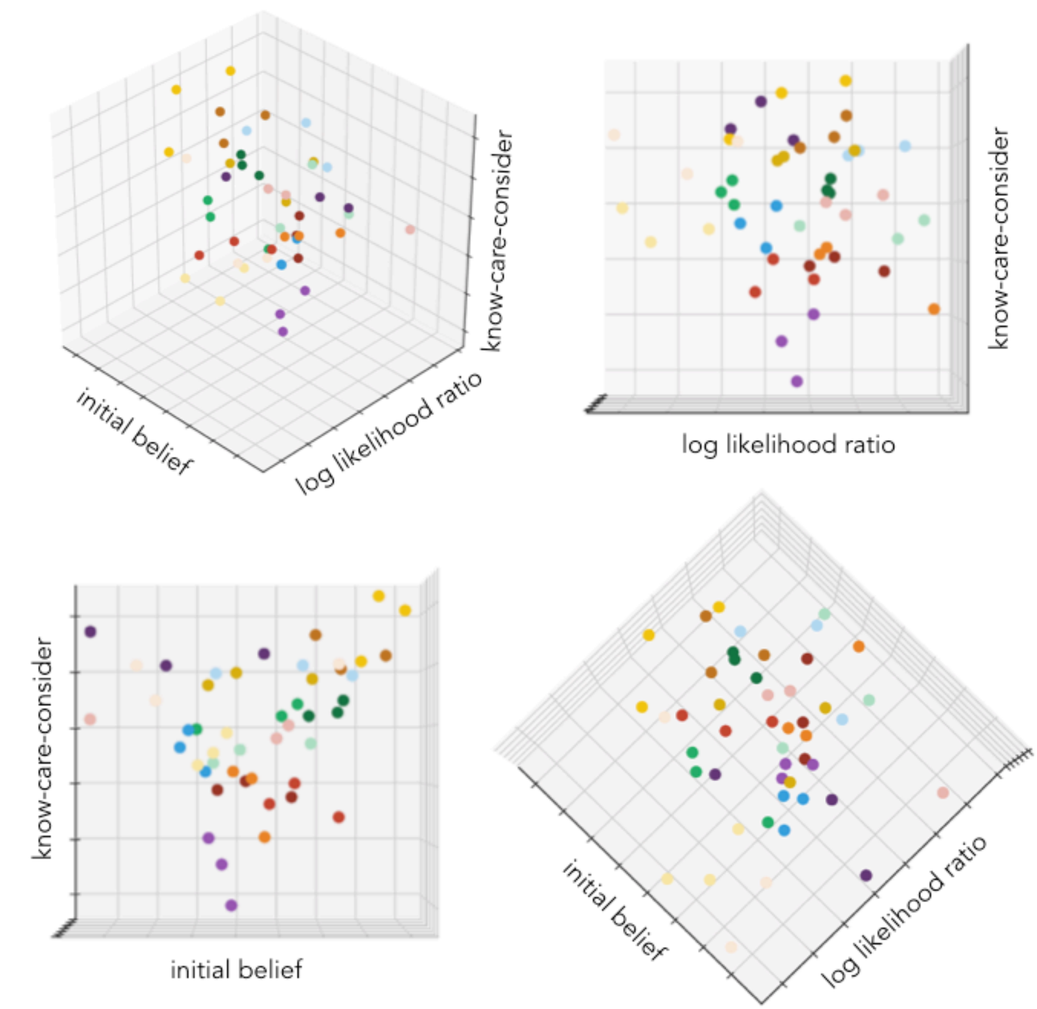

Figure 26: Sorting of ballot measures based on initial belief, persuasiveness (log likelihood ratio), and the first principal component of the knowledge, consideration, and care ratings. Each color represents one cluster. The three ballot measures in each cluster were randomly assigned to the three stimuli sets.

News Headlines: Participants were shown newspaper headlines and had to judge whether each headline was real or fake. When judging whether a headline was real or fake, participants were told to consider how likely the event described in the headline actually took place in real life. In each trial, a newspaper headline was presented and participants used a digital slider to report their response to the question "How likely did this ACTUALLY happen?" (Figure 27). The leftmost point on the slider was labeled "definitely fake" and the rightmost point was labeled "definitely real." Subtle vertical lines were drawn at 25%, 50%, and 75% of the slider track to help orient participants.

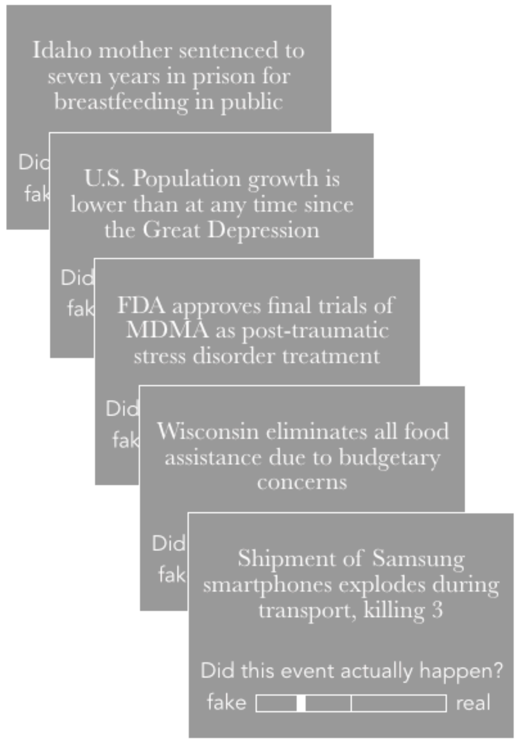

Figure 27: Example news headlines and task structure.

Sixty real news headlines were pulled from the Washington Post, the New York Times, Fox News, BBC, and NPR and were published between October 2016 and July 2017. Sixty fake headlines were pulled from Snopes, an online fact-checking website. The fake headlines had to have been published on a fake news site or social media site and had to have been rated "false" by Snopes. All but 2 of the 60 fake headlines were published in 2016 or 2017. Each headline was assigned a political bias score, ranging from -3 for very liberal (for example, "Mike Pence opposes word 'vice' on religious grounds, doesn't want to be called vice president"), to 0 for neutral ("Man dressed as clown arrested after police find 11 bodies stuffed in freezers at his home"), to 3 for very conservative ("Van full of illegal immigrants caught visiting Arizona voting booths to vote for

Clinton"). A separate group of 59 participants rated all 120 news headlines using the digital slider. The headlines were sorted into three stimuli sets on the basis of believability (estimated by the mean slider position of each headline) and political bias. The real and fake headlines were each sorted into 20 groups of 3 headlines each using the same iterative agglomerative clustering method that was used to sort the ballot measures (Figure 28). Then, the headlines in each cluster were randomly assigned to the three stimuli sets. The headlines in each stimuli set did not significantly differ in believability, political bias, response time, or accuracy.

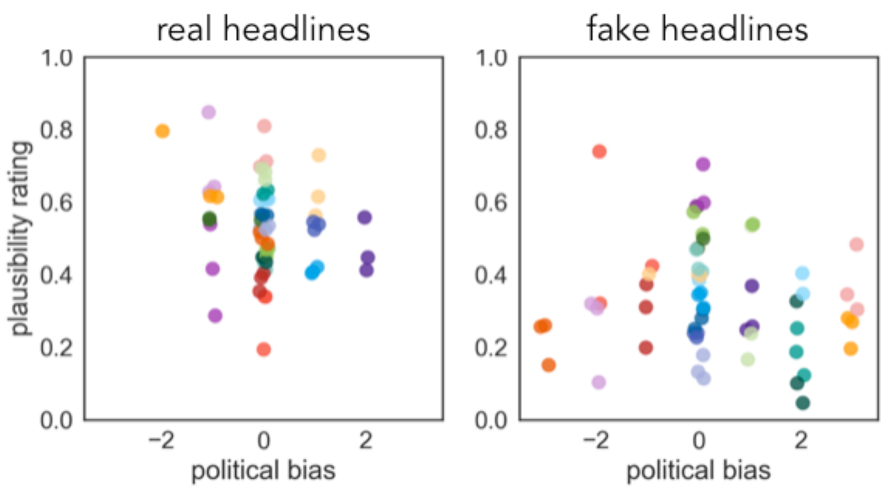

Figure 28: Sorting of real and fake headlines based on plausibility and political bias. Each color represents one cluster. The three headlines in each cluster were randomly assigned to the three stimuli sets.

3. Current Flow Modeling

The current flow modeling pipeline was identical to that used in the first tDCS experiment (see section II).

3. Stimulation Procedure

Each participant completed three sessions: a LH-bias condition with a left anode and right cathode, a RH-bias condition with a right anode and left cathode, and a sham condition which served as a control. The stimulation sessions were completed at the same time of day when possible and at least 48 hours apart.

Participants signed a consent form and read task instructions at the beginning of the first session. At the beginning of all sessions, participants reported their baseline opinions of the ballot measures using the digital slider. Next, participants were fitted with the electrodes using the exact same procedure described in chapter II. This experiment deviated from the first tDCS experiment in only two ways: first, the anode was set to +1.5mA and the cathode was set to -1.5mA instead of +2mA and -2mA, and, second, participants were stimulated for 35 minutes instead of 20 minutes.

During stimulation, participants completed 16 trials of the voter ballot task, 16 trials of the criminal court task, and 40 trials of the news headlines task. After the third session, participants completed an online survey that consisted of questions about motivation (McAuley, Duncan, & Tammen, 1989), handedness (Oldfield, 1971), political attitudes and belief superiority (Toner, Leary, Asher, & Jongman-Sereno, 2013), dogmatism (Altemeyer, 2002), political attitudes (Pew Research Center, Political Polarization in the American Public: Appendix A), and news consumption (adapted from Pew Research Center, Americans' Attitudes about the News Media Deeply Divided along Partisan Lines: Appendix A).

4. Data Analysis

All analyses were performed in Python using the Pandas, Numpy, Sci-kit learn, and statsmodels packages.

To determine how prior beliefs and evidence affect participants' belief updates, continuous heatmaps were created that denoted participants' average belief update over the full range of prior belief certainties (from completely uncertain to completely certain) and evidence consistency (from highly confirmatory to highly disconfirmatory of the prior). The magnitudes of the belief updates were equal to the absolute difference in slider position before vs. after each observation and the directions of the update were normalized so that positive values indicated updates toward either end of the slider (advances) and negative values indicated updates toward the middle of the slider (backtracks). Prior certainties ranged from 0 to 0.5. Evidence consistency was estimated with the log likelihood ratios associated with each piece of evidence from the respective pilot experiments. Positive values indicated that the evidence was aligned with participants' prior beliefs and vice versa. Continuous heatmaps were generated by summing Gaussian kernels with heights equal to the belief update, x coordinates equal to the prior belief certainty, and y coordinate equal to the evidence consistency for all observations. Heatmaps were subtracted from each other to compare stimulation conditions.

Multinomial logistic regression was used to determine the effects of stimulation on the frequency of different types of belief updates (advances, backtracks, or no updates) for both the criminal court cases and ballot measures. The design matrix for both analyses consisted of the following regressors: LH-bias stimulation; RH-bias stimulation; prior belief certainty (continuous), evidence inconsistency (continuous); LH*prior certainty; RH*prior certainty; LH*evidence inconsistency; RH*evidence inconsistency; LH*prior certainty*evidence inconsistency; RH*prior certainty*evidence inconsistency; evidence number; evidence order; degree of handedness (continuous); dogmatism score (continuous); self-reported effort (continuous); self-reported skill (continuous); belief superiority score (continuous), and political attitude (continuous: more negative, more liberal).Continuous regressors were normalized so that they ranged from 0 to 1. Beta values were found with Ordinary Least Squares and a Bonferroni correction was applied to correct for multiple comparisons (alpha = 0.0023).

C. Results

1. Criminal court case scenarios

Pilot data: A pilot study was conducted on a separate group of 48 participants in order to characterize the court case scenarios. Each participant provided judgments for all 48 scenarios and the evidence order was counterbalanced across participants.

Participants were given an introduction to a court case and asked to judge whether the defendant was guilty or not guilty before seeing any evidence. For this first judgment, most participants kept the slider in the middle of the track, but 3 participants consistently moved the slider all the way to "definitely not guilty" and 3 participants moved the slider closer toward "guilty." The persuasiveness of evidence in each court case was estimated by the log likelihood ratio (Equation 2) where positive values indicate that participants were more likely to make a "guilty" verdict after seeing all the evidence and negative values indicate that participants were more likely to make a "not guilty" judgment. The 48 scenarios are sorted from most to least persuasive of guilt in Figure 29. Although all of the cases resulted in convictions that were later overturned, about half of the cases were associated with greater "guilty" verdicts and about half were associated with greater "not guilty" verdicts.

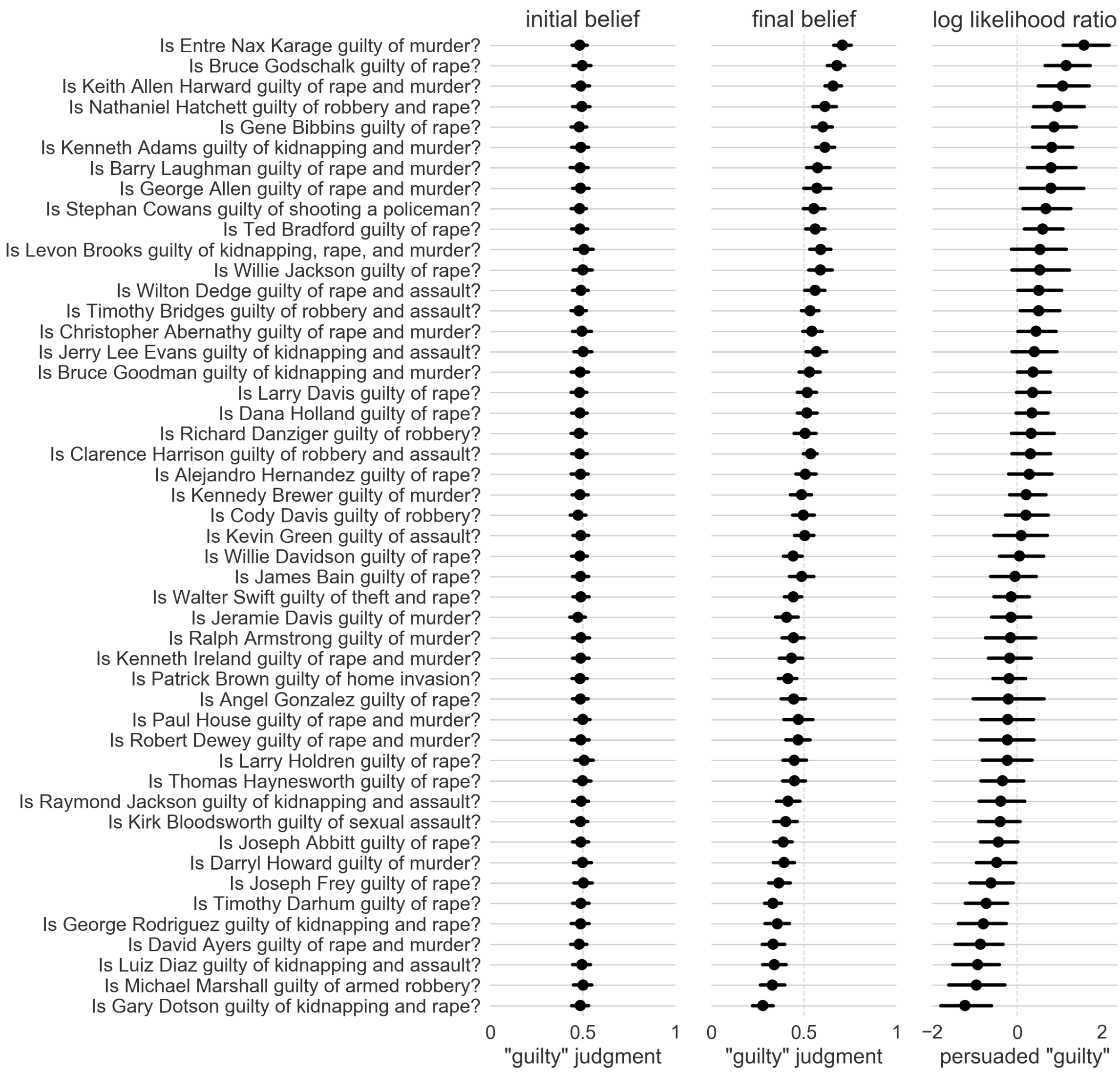

Figure 29: Pilot ratings for criminal court case scenarios. Average initial belief, final belief, and persuasiveness associated with 48 criminal court case scenarios. Each bar represents the mean of 48 participants and error bars represent 95% confidence intervals.

Crime scene evidence was sorted into 22 categories prior to the pilot experiment. For each evidence presentation, the log likelihood ratio associated with that piece of evidence was computed (using Equation 2) and the log likelihood ratios for each evidence category were averaged in order to determine which types of evidence were the most persuasive. We found that participants were most strongly persuaded that a defendant was guilty when presented with confessions, matches between physical crime scene evidence and the suspect, and second-hand confessions (i.e., when a third party reported that the suspect confessed). Participants were less persuaded by connections to the victim and prior criminal records. Physical evidence mismatches, a lack of physical evidence, and suspect description mismatches were the most convincing of a defendant's innocence. There was a striking asymmetry in how participants weighed suspects' statements – based off of participants' ratings, confessions were approximately 13x more indicative of guilt than innocence, while claims of innocence were only 1.4x more indicative of innocence than guilt.

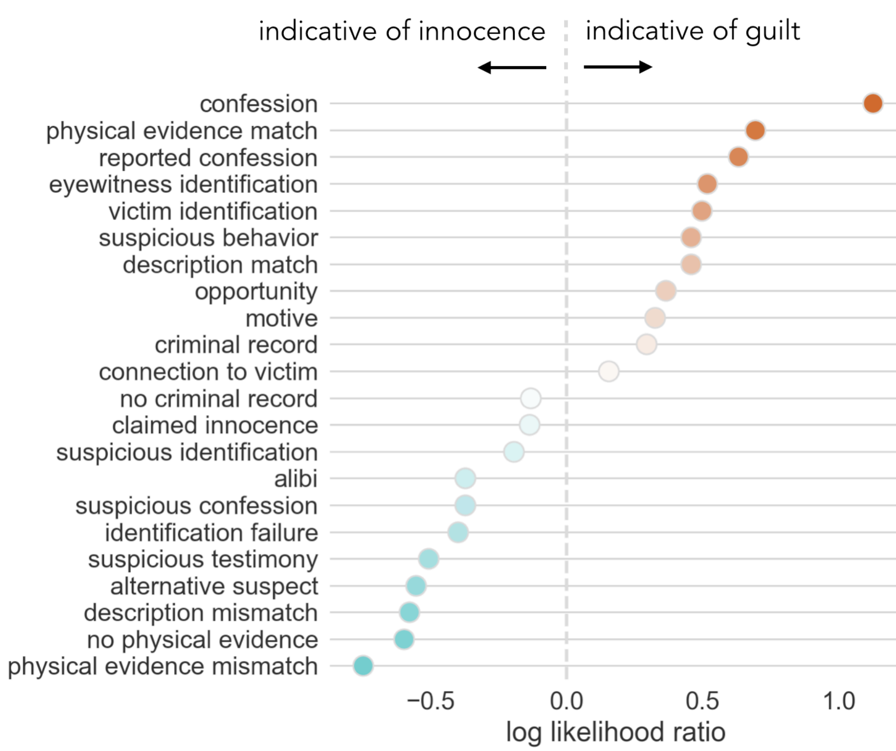

Figure 30: Average log likelihood ratio for each evidence category. Positive values indicate the evidence is more likely to occur when a defendant is guilty and negative values indicate that the evidence is more likely to occur when a defendant is innocent.

tDCS data: Participants were presented with evidence from real criminal investigations and were asked to judge whether the defendant was guilty or not guilty. For each of the 48 criminal cases, the evidence was presented in one of four orders that differed by presentation sequence (guilty evidence last vs. innocent evidence last) and order type (stacked evidence types vs. interleaved evidence types). The same pieces of evidence were presented for each court case regardless of evidence order. Figure 31 shows participants' judgments in every trial with the mean judgments for each evidence order overlaid in black. On average, participants moved the slider toward "guilty" after receiving guilty evidence (represented with orange bars) and closer toward "not guilty" after receiving innocent evidence (blue bars). Most participants placed the slider in the middle of the track after being introduced to the crime, but before seeing any evidence. However, one participant always moved the slider all the way to "not guilty" and one participant moved the slider different distances toward "guilty". Participants' final judgments varied greatly, which is likely due to the strength of each court case's body of evidence. There was a slight primacy effect for stacked evidence orders: after all evidence was presented, the same defendants were judged to be less guilty when innocent evidence was presented first than when guilty evidence was presented first (0.467 vs. 0.497; sufficient-summary-statistic with nested scenario effects, $z=-2.008$, $p=0.044$). There was no difference in final beliefs when the evidence types were interleaved (guilty last: 0.471 vs. innocent last: 0.478; sufficient-summary-statistic with nested scenario effects, $z=-0.491$, $p=0.624$).

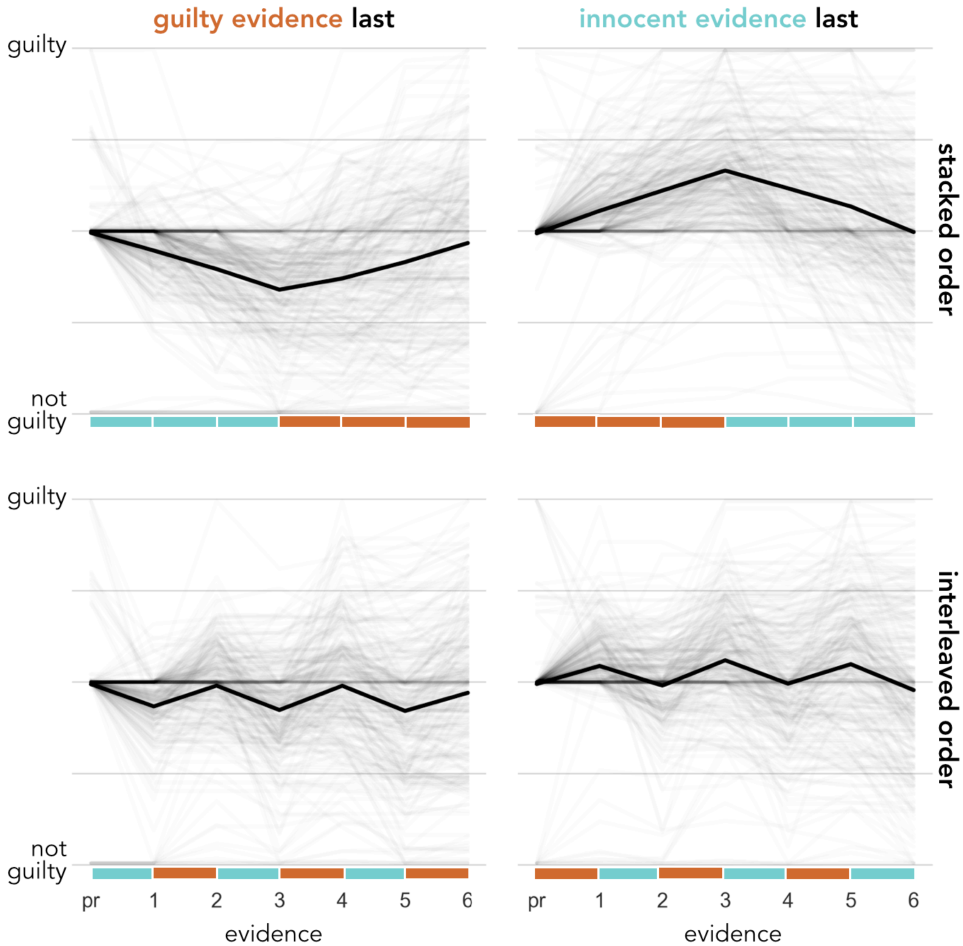

Figure 31: Beliefs over time for criminal court cases. The evidence for all 48 scenarios was presented in 4 different ways, with each evidence order being presented to 6 participants each. The evidence orders varied according to order type (stacked or interleaved) and presentation sequence (guilty evidence last or innocent evidence last). Bars indicate when guilty (orange) or innocent (blue) evidence was presented. Time courses for individual trials are in light gray and mean time courses are in black.

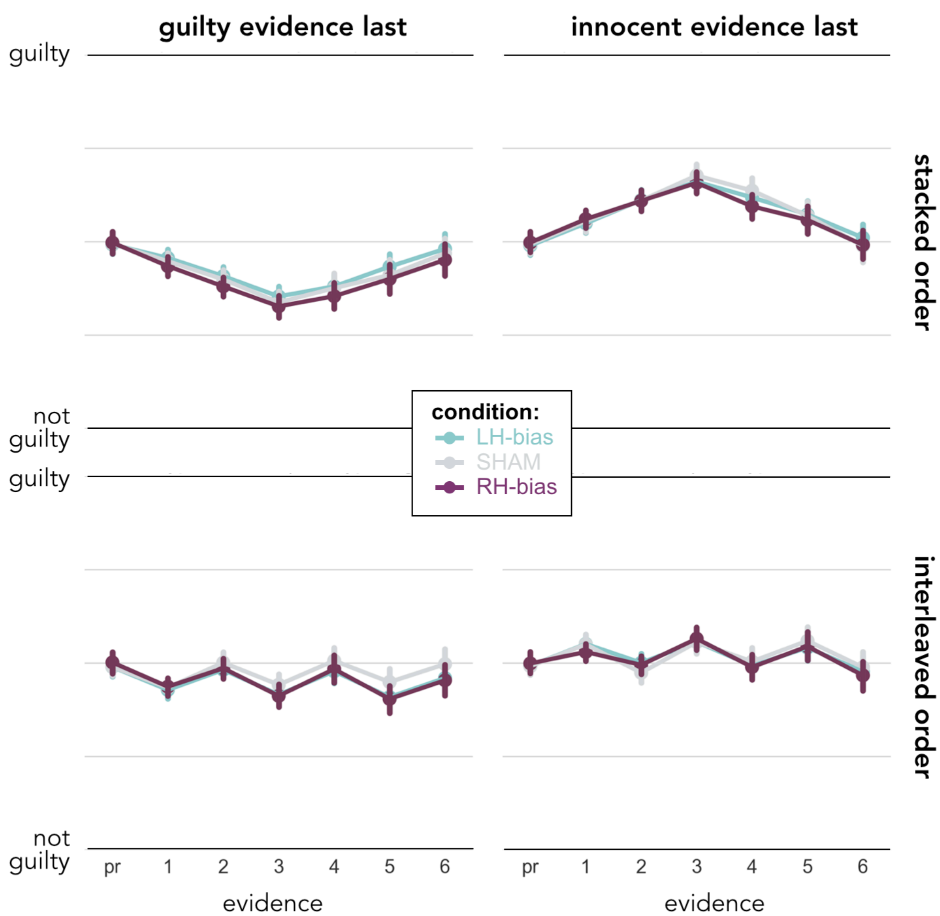

Figure 32: Beliefs over time according to stimulation condition. The mean time courses for each evidence order type under LH-bias stimulation (blue), RH-bias stimulation (purple), or sham stimulation (gray). Error bars represent the standard error of participant means.

Belief trajectories were very similar across the three stimulation conditions for both stacked and interleaved presentation orders (Figure 32). The rationale for using stacked evidence orders was to spur the formation of strong beliefs and then introduce inconsistent evidence to create conflict. These trials allow for two tests of our framework. First, since uncertainty is high at the beginning of each trial as details about the crime emerge, we expected that LH-bias stimulation would be associated with greater belief updates relative to sham and RH-bias stimulation for the first three pieces of evidence (which either all suggested innocence or all suggested guilt). Second, we predicted that the first inconsistent piece of evidence (evidence 4) would produce greater belief backtracks in the RH-bias stimulation condition compared to LH-bias and sham stimulation. To test these predictions, we first reversed the slider position coding of trials in which guilty evidence was presented last so that, for all stacked trials, increases in slider position indicated greater agreement with initial evidence. Figure 33 shows the changes in beliefs for the first three pieces of evidence relative to the prior belief for LH-bias stimulation (blue), sham stimulation (gray), and RH-bias stimulation (purple). Contrary to predictions, LH-bias stimulation was associated with numerically smaller belief updates in the direction of the evidence compared to RH-bias stimulation and sham, although none of these differences reached significance using the sufficient-summary-statistic with nested scenario effects. In line with predictions, RH-bias stimulation was associated with greater backtracks and LH-bias stimulation was associated with smaller backtracks compared to sham for the 4th piece of evidence, however, these differences were also not significant.

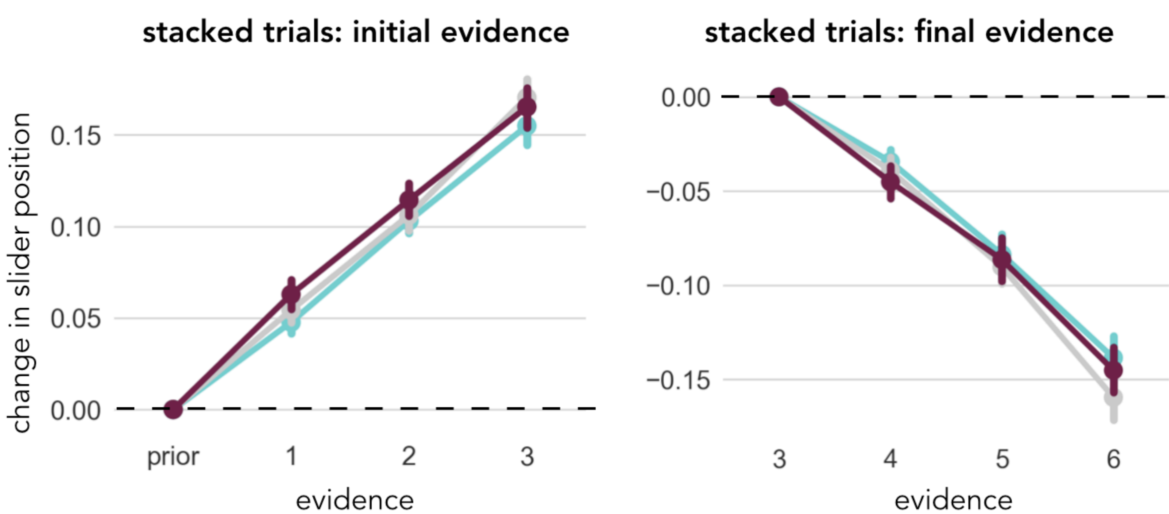

Figure 33: Belief changes in trials with stacked evidence. Left: Changes in slider position for the first three pieces of evidence relative to the slider position for the introduction. Right: Changes in slider position for the last three pieces of evidence relative to the slider position for third piece of evidence. Circles indicate mean slider positions for each evidence number and each condition (blue: LH-bias, gray: sham, purple: RH-bias). Greater values indicate slider positions that are closer to the initial evidence type. Error bars represent 95% confidence intervals of scenario means.

The previous analysis took advantage of the structure of the evidence presentation orders, but made assumptions about belief strength and the presence of belief-evidence conflicts. As a more direct test of our hypotheses, we used data from all evidence presentations and characterized how stimulation influenced belief updating across the full spectrum of prior belief certainties and evidence consistencies. Our central prediction concerned belief updating when evidence disconfirmed certain beliefs: we predicted that LH-bias stimulation would be associated with less backtracking compared to sham stimulation since backtracking increases uncertainty, and that RH-bias stimulation would be associated with more backtracking compared to sham since backtracking reduces conflict. Consistent with predictions, we found that RH-stimulation was associated with greater backtracking when evidence disconfirmed strong prior beliefs (approximately, when certainty > 0.4) compared to both sham stimulation and LH-bias stimulation.

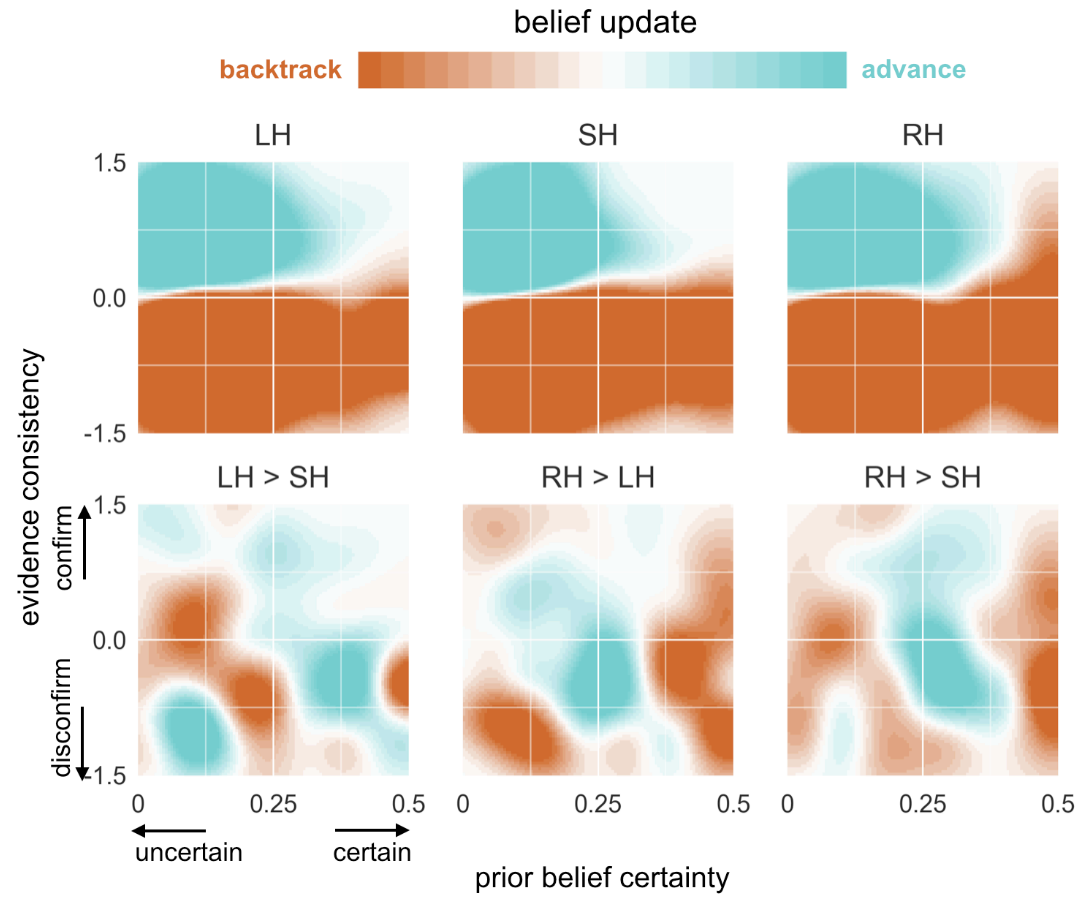

Figure 34: Normalized belief updates according to prior certainty and evidence consistency for criminal court cases. Areas in red denote backtracks in beliefs (updates toward the middle of the slider) and areas in blue denote advances in beliefs (updates toward either end of the slider). Evidence consistency is estimated from the log likelihood ratios computed from the pilot experiment.

To examine the effects of additional factors on belief updating, we ran a multinomial logistic regression analysis. First, we classified all observations into three belief update types: 1) backtracks, defined as movements toward the middle of the slider track, (n=2112), 2) advances, defined as movements toward either end of the track (n=3314), and 3) non-updates, in which the slider was not moved at all (n=1774). The design matrix included interactions for stimulation condition, prior belief certainty, and evidence consistency, as well as factors of non-interest including evidence order, evidence number, and participant factors (handedness, self-reported effort, self-reported skill, belief superiority, dogmatism, and political attitude). Cortical current density was not included in the model due to the poor interpretability of multiple 4-way interactions. Multinomial logistic regression was used to determine the marginal influence of each factor on the likelihood of each update type. The results of the analysis are presented in Figure 35, and can be interpreted as, for example, "disconfirmatory evidence makes it more likely that an update will be a backtrack, less likely that an update will be an advance, and does not change the likelihood of keeping the slider in place." Parameters that are significantly different from zero after controlling for multiple comparisons are outlined in black.

As expected, inconsistent evidence (that went against participants' prior beliefs) increased the likelihood of backtracks and decreased the likelihood of advances. Certain prior beliefs were associated with a lower likelihood of advancing the slider and a greater likelihood of backtracking the slider or keeping it in place. Conflict was associated with fewer backtracks. Although this seems unintuitive, it makes sense considering that conflict is defined as the interaction between evidence consistency and prior certainty. There are more backtracks as evidence is more disconfirmatory and as prior certainty increases, but comparatively fewer backtracks when both occur together. There was no effect of stimulation on belief update type, and no significant interactions between stimulation, evidence consistency, and prior certainty on belief updating.

Some of the participant measures were significant, including belief superiority, self-reported effort on the task, political conservatism, right-handedness, and dogmatism. Of note, higher belief superiority and political conservatism (or in this sample, less political liberalism) were associated with fewer backtracks. Greater dogmatism was associated with more advances, while self-reports of greater effort on the task and more extreme right-handedness was associated with fewer advances.

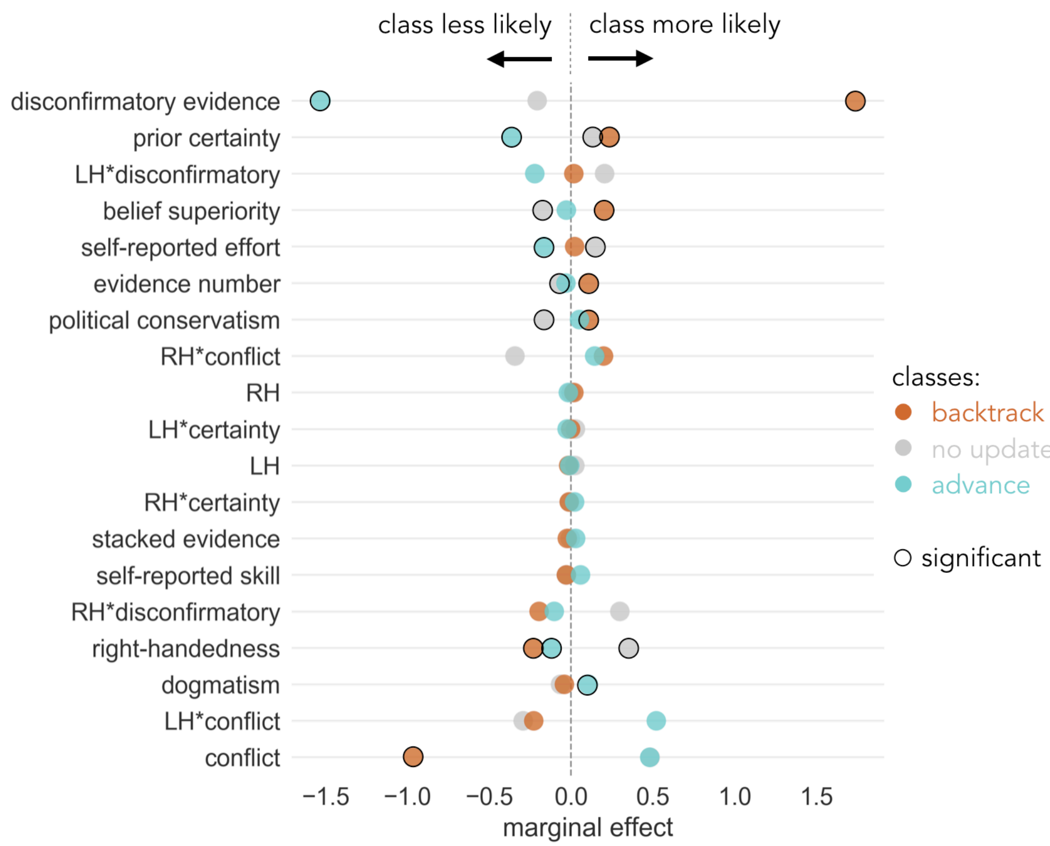

Figure 35: Marginal effects of multinomial logistic regression of update types for the criminal court cases. Marginal effects of each factor on the likelihood that a given observation is a backtrack (orange), advance (blue), or no update (gray). Positive and negative marginal effects indicate that the factor increases and decreases the likelihood of the class type, respectively. Significant effects are outlined in black (Bonferroni correction for 19 comparisons, $t>3.01$, $p<0.0026$).

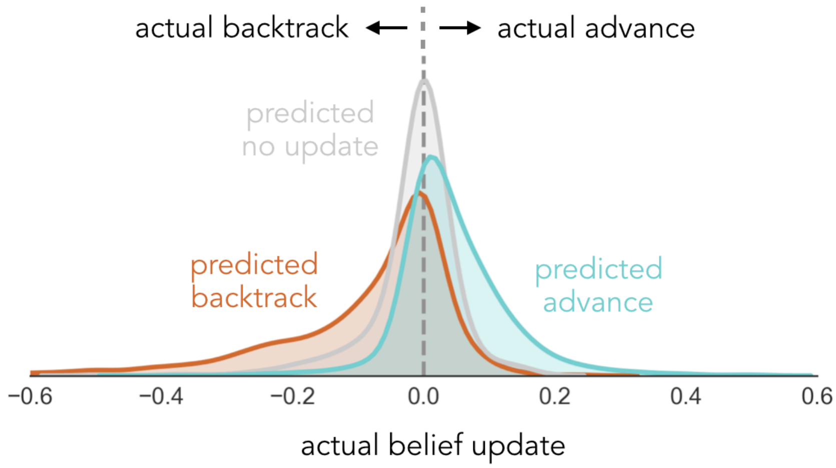

Figure 36: Predictions of the multinomial regression analysis for court cases. Distributions of actual normalized updates for predicted backtracks (orange), advances (blue), and non-updates (gray). True advances are defined as positive normalized updates and true backtracks are defined as negative normalized updates.

2. Ballot Measure Scenarios

Pilot data: Like the criminal court cases, the ballot measures were also characterized by an independent group of 48 participants. Each participant provided judgments for all 48 scenarios and the evidence order was counterbalanced across participants.

As in the court case pilot study, participants were given an introduction to the ballot measure and asked to report their initial opinion before seeing any arguments. Participants were also asked how much they knew about the topic, how much they cared about it, and how much they had considered the issue beforehand. Participants' initial beliefs varied greatly from topic to topic. In Figure 37, ballot measures are sorted form strongest initial "yes" votes to strongest initial "no" votes.

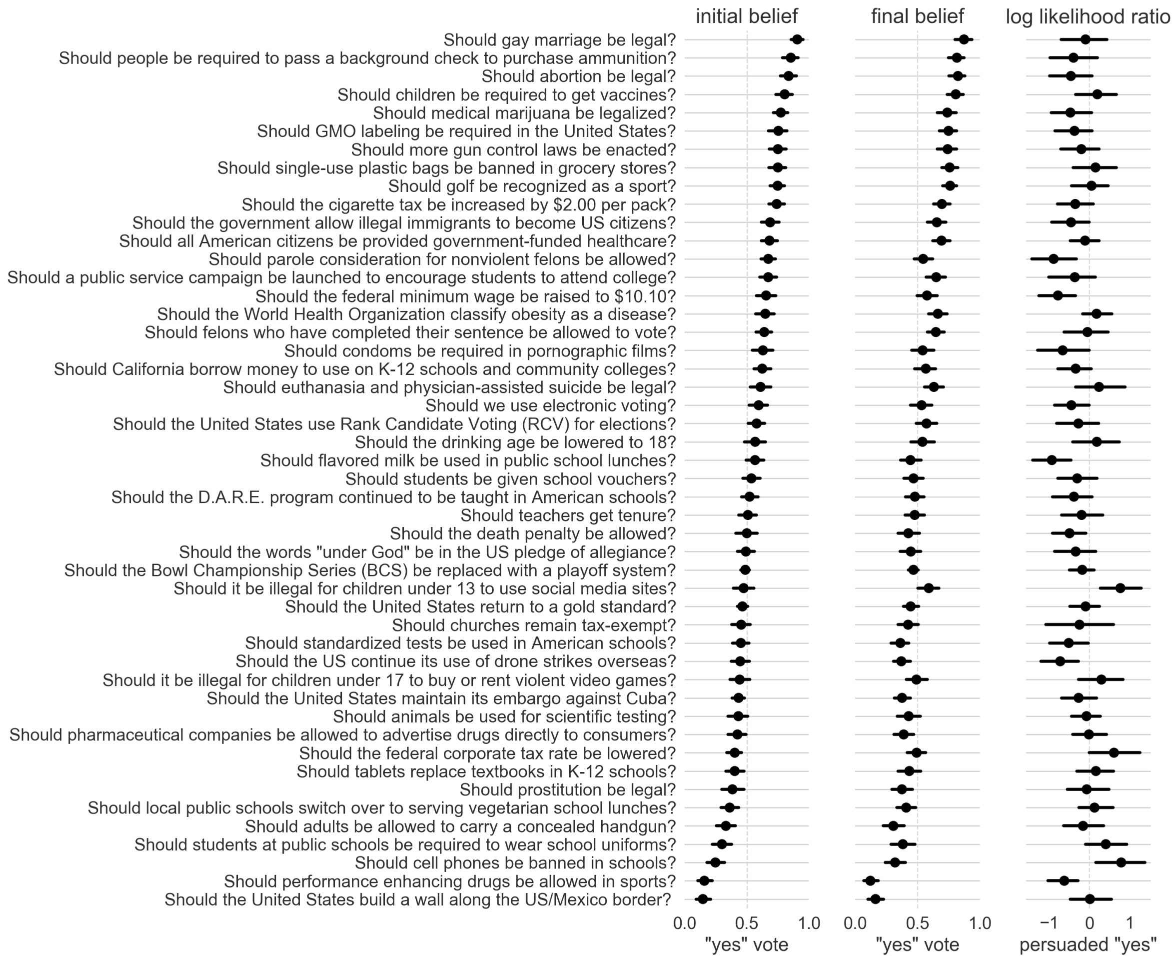

Figure 37: Pilot ratings for ballot measures. Average initial belief, final belief, and persuasiveness associated with 48 ballot measures. Each bar represents the mean of 48 participants and error bars represent 95% confidence intervals.

tDCS data: Participants were presented with pro and con arguments for hypothetical ballot measures and were asked to vote yes or no on the measure. Each of the 48 ballot measures were paired with two arguments in favor of the measure and two arguments in opposition to it and the argument orders differed according to which argument was presented last and whether the argument types were stacked or interleaved. The same arguments were presented for each ballot measure regardless of evidence order. Participants' opinions over the course of every trial are presented in Figure 38, with the mean judgments for each evidence order overlaid in black. Participants generally moved the slider toward "yes" after reading a pro argument (represented with orange bars) and closer toward "no" after reading con arguments (blue bars). There are several striking differences between the individual trials in the ballot measures task vs. the criminal cases task. First, there is much more variation in participants' prior beliefs for the ballot measures than for the criminal cases. Second, as evinced by the darker lines at the "yes" and "no" boundaries, participants occasionally moved the slider all the way to one end of the slider and kept it there throughout all of the argument presentations.

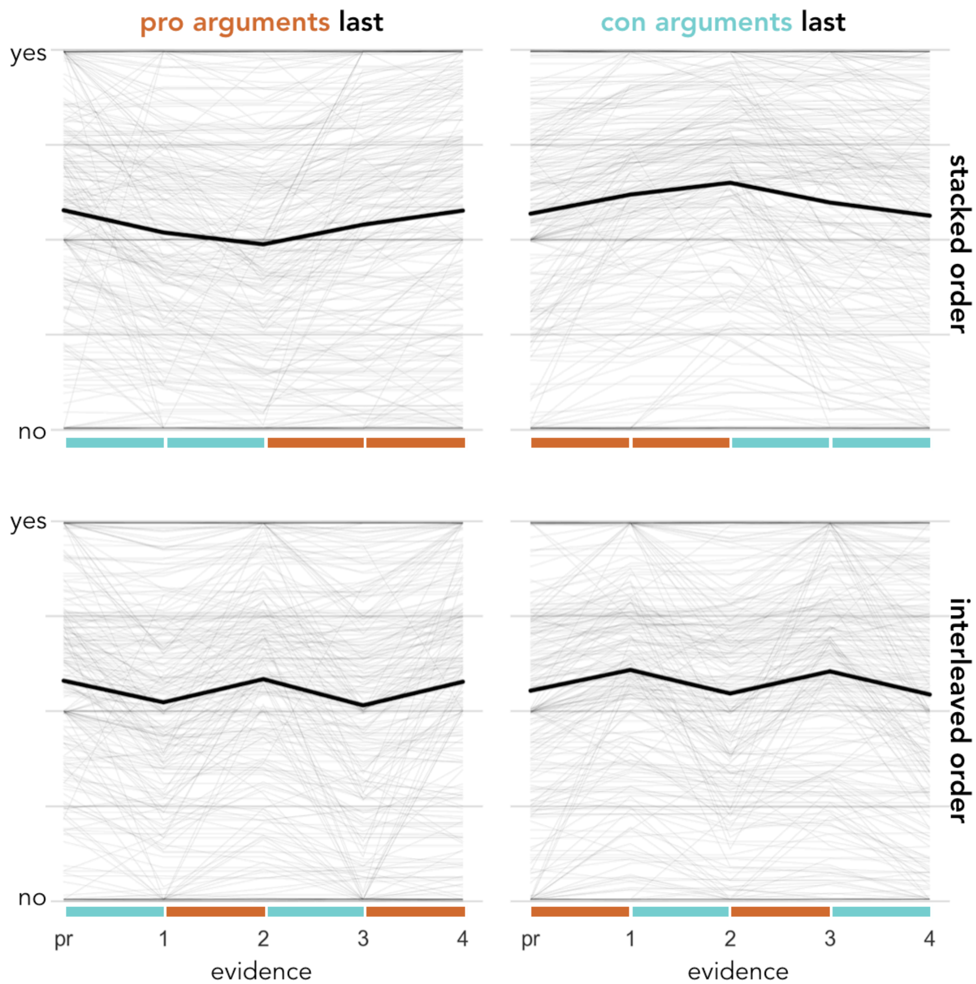

Figure 38: Beliefs over time for ballot measures. The arguments for all 48 scenarios were presented in 4 different orders, with each evidence order being presented to 6 participants each. The evidence orders varied according to order type (stacked or interleaved) and presentation sequence (pro argument last or con argument last). Bars indicate when pro (orange) or con (blue) arguments were presented. Time courses for individual trials are in light gray and mean time courses are in black.

To illustrate how greatly baseline beliefs differ across different ballot measures, belief ratings for each ballot measure were averaged to create a mean belief trajectory for each ballot measure and each evidence presentation order (Figure 39). Since ballot measures were associated with a wide range of baseline beliefs, subsequent analyses used baseline subtracted slider positions (by subtracting the slider position for the introductory text) and nested statsistical tests were performed across ballot measures rather than across participants.

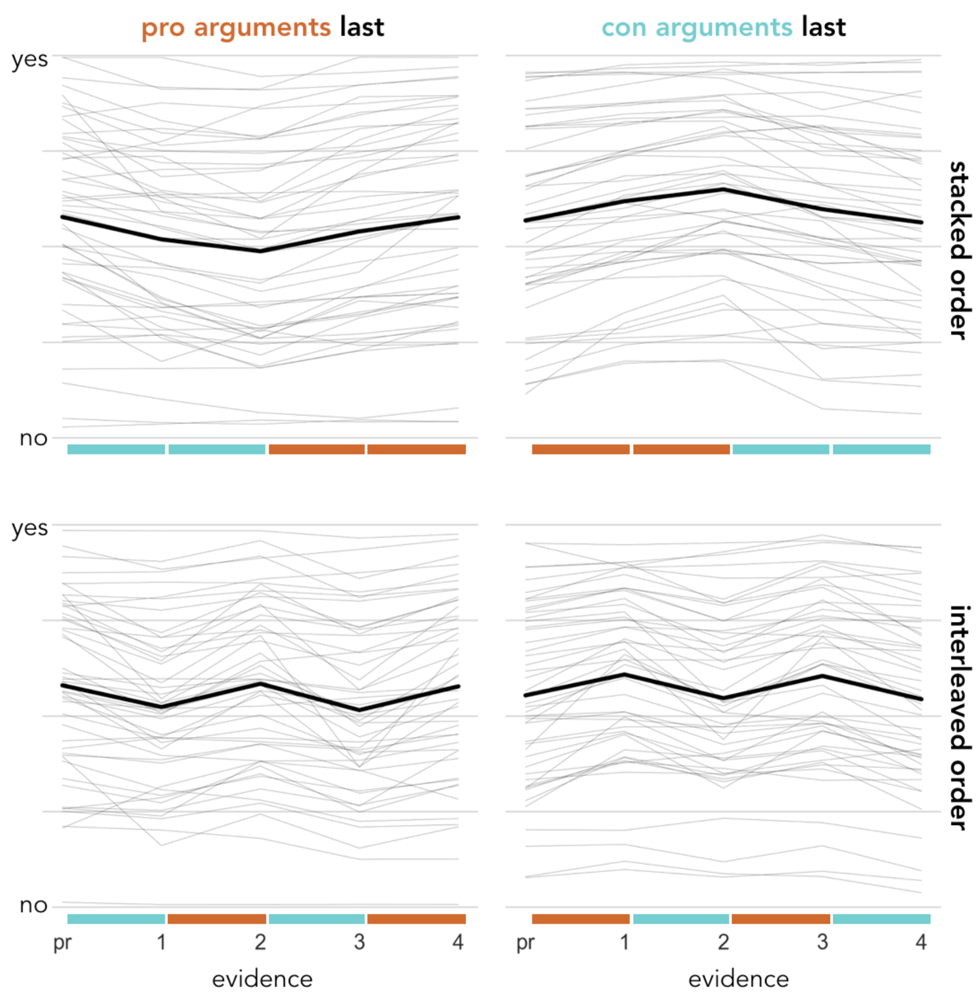

Figure 39: Average beliefs over time for each ballot measure. Each light gray line represents the average belief time course for one ballot measure. The means were computed across 6 participants for each evidence order.

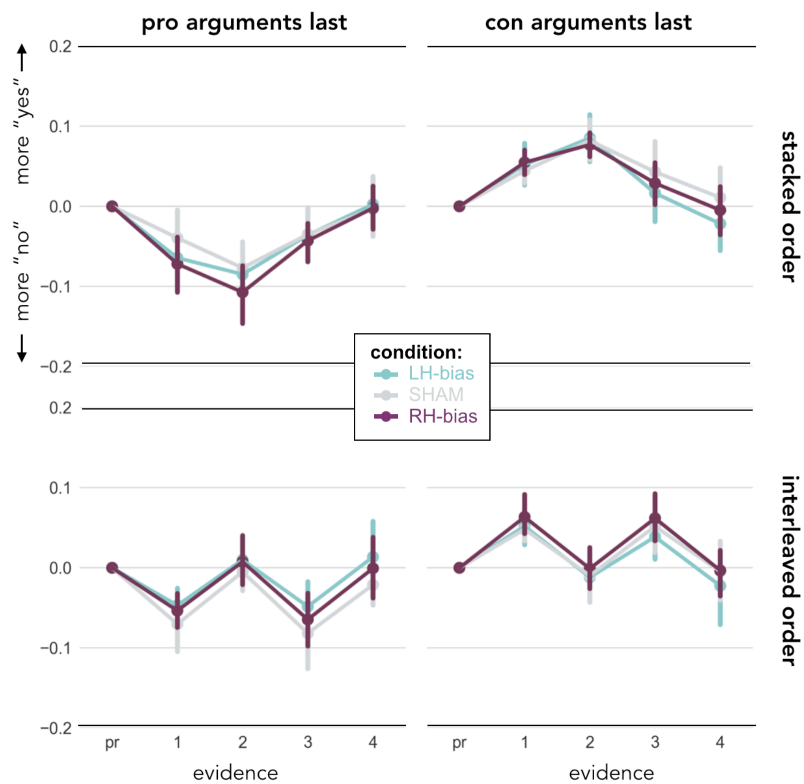

Figure 40: Beliefs on ballot measures in each stimulation condition. The mean time courses for each evidence order type under LH-bias stimulation (blue), RH-bias stimulation (purple), or sham stimulation (gray). Error bars represent the standard error of participant means.

Average belief trajectories for the three stimulation conditions are presented in Figure 40. As before, we examined how tDCS influenced belief updating across the full spectrum of prior belief certainties and evidence consistencies. Again, we predicted to see the largest condition differences when evidence disconfirmed certain beliefs, such that LH-bias stimulation would lead to less backtracking relative to sham and RH-bias stimulation would lead to greater backtracking relative to sham. We found that RH-bias stimulation was generally associated with greater backtracking when evidence disconfirmed strong prior beliefs compared to both sham stimulation and LH-bias stimulation (Figure 41).

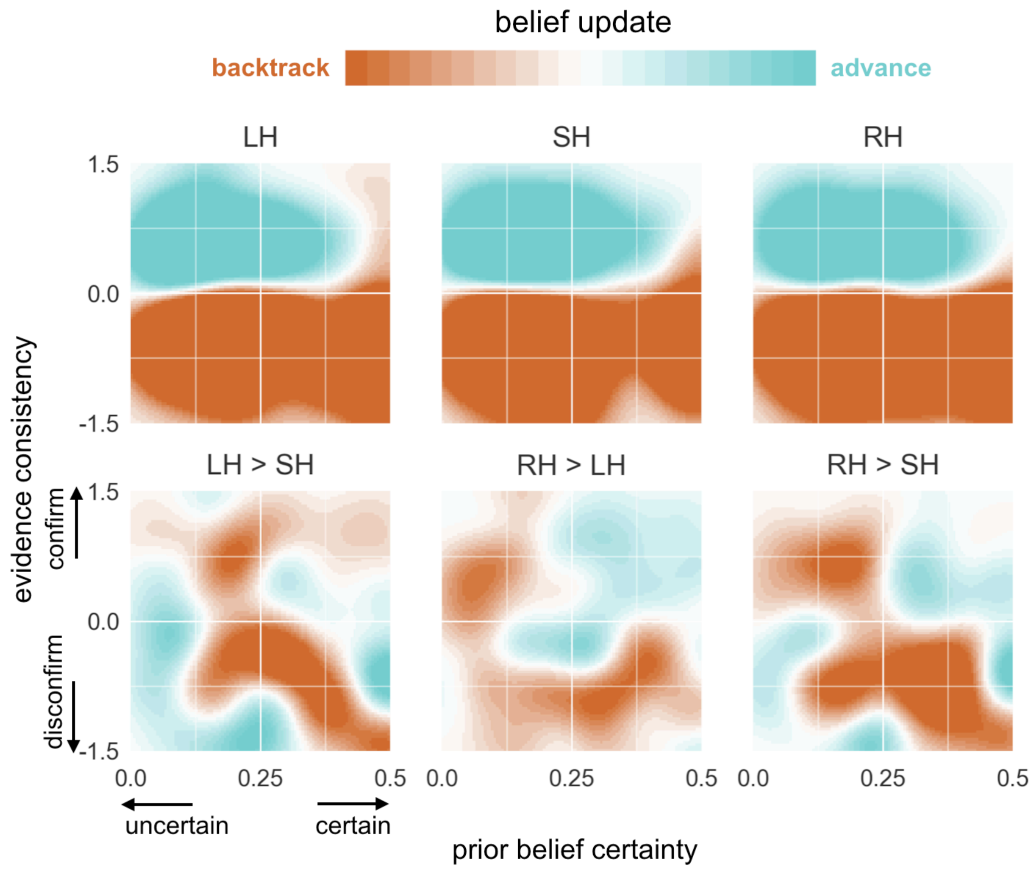

Figure 41: Normalized belief updates according to prior certainty and evidence consistency for ballot measures. Areas in red denote backtracks in beliefs (updates toward the middle of the slider) and areas in blue denote advances in beliefs (updates toward either end of the slider). Evidence consistency is estimated from the log likelihood ratios computed from the pilot experiment.

Finally, we used multinomial logistic regression to examine which factors were associated with backtracks, advances, and no belief updates. All observations were classified as backtracks (n=1437), advances (n=1443), or non-updates (n=1895). The design matrix used in the criminal court case analysis was used in this analysis as well. Multinomial logistic regression was used to determine the marginal influence of each factor on the likelihood of each update type (Figure 42). Parameters that are significantly different from zero after controlling for multiple comparisons are outlined in black.

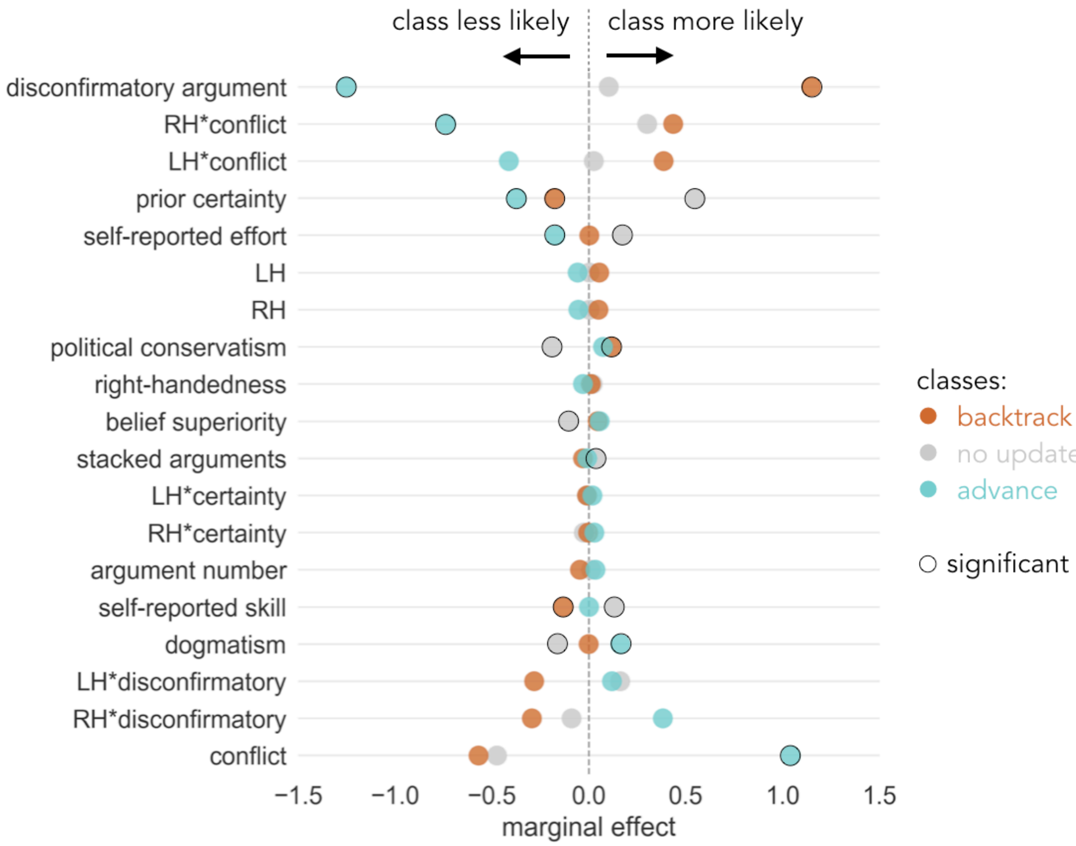

Figure 42: Marginal effects of multinomial logistic regression of update types for the ballot measures. Marginal effects of each factor on the likelihood that a given observation is a backtrack (orange), advance (blue), or no update (gray). Positive and negative marginal effects indicate that the factor increases and decreases the likelihood of the class type, respectively. Significant effects are outlined in black (Bonferroni correction for 21 comparisons, $t>3.01$, $p<0.0026$).

As expected, disconfirmatory arguments (that went against participants' prior beliefs) increased the likelihood of backtracks and decreased the likelihood of advances. Certain prior beliefs were associated with a lower likelihood of moving the slider (either backtracking or advancing) and a greater likelihood of keeping the slider in place. Conflict (defined as the interaction between prior certainty and disconfirmatory evidence) was associated with a greater likelihood of advance.

The three-way interaction between RH-bias stimulation, prior belief certainty, and evidence disconfirmation (RH-bias*conflict) was associated with a lower likelihood of advance, meaning that participants were less likely to adopt more extreme beliefs in the face of conflict under RH-bias stimulation compared to sham stimulation. Similar to the court case analysis, several participant factors influence belief update type, including self-reported effort, political conservatism, belief superiority, self-reported skill, and dogmatism.

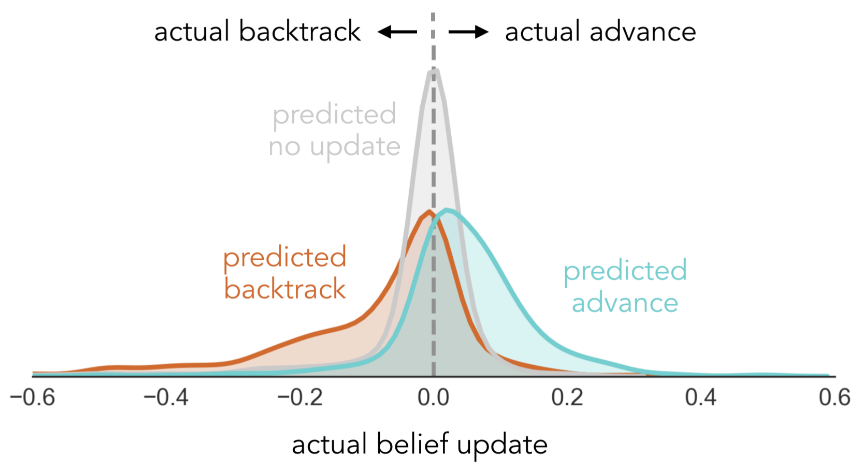

Figure 43: Predictions of the multinomial regression analysis for ballot measures.

3. News Headlines

In each stimulation session, participants were shown 40 news headlines, 20 of which were real and 20 of which were fake, and had to judge the likelihood that each headline was real. We predicted that RH-bias stimulation would make participants more skeptical of headlines in general.

In order to compute hit rates and false alarm rates, participants' plausibility judgments were binarized such that likelihoods < 0.5 were considered "fake" judgments and likelihoods > 0.5 were considered "real" judgments. Likelihoods exactly equal to 0.5 were excluded from the analysis and made up 1.2% of all responses. Participants' discrimination of real vs. fake headlines (d') and response criterion (c) were computed for each stimulation condition (Figure 44 and Figure 45). Compared to LH-bias stimulation and sham, RH-stimulation was associated with poorer discrimination of real vs. fake headlines (LH>RH Wilcoxon signed-rank test, statistic=74.5, p=0.03, RH>SH Wilcoxon signed-rank test, statistic=68.0, p=0.03). However, there was no significant difference in response criterion between LH-bias and RH-bias stimulation, although participants were slightly more conservative under RH-bias stimulation (LH>RH Wilcoxon signed-rank test, statistic=139.0, p=0.75, RH>SH Wilcoxon signed-rank test, statistic=111.5, p=0.27).

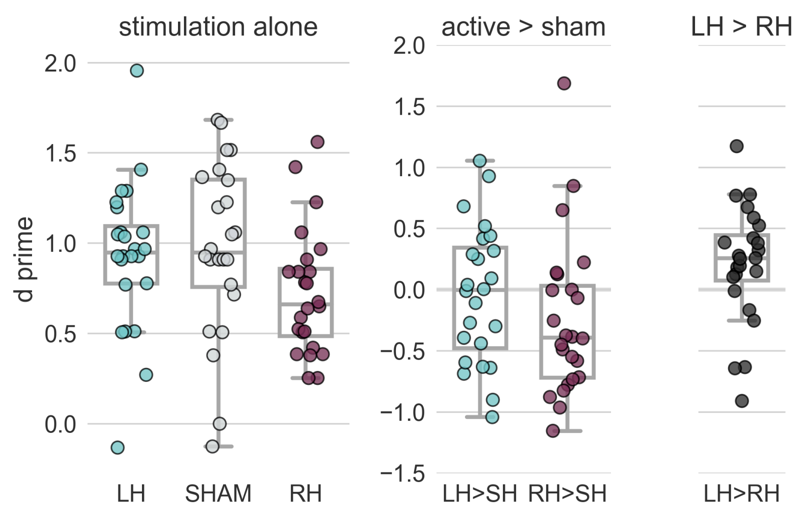

Figure 44: Condition differences in discrimination ability (d'). D' values of 0.5 indicate chance performance and higher values represent better discrimination.

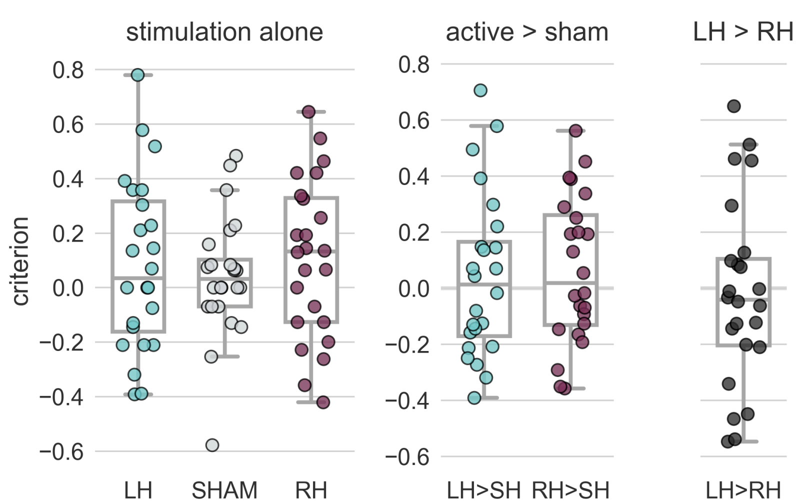

Figure 45: Condition differences in response criterion. Positive values correspond to conservative criteria and negative values correspond to liberal criteria.

Participants' poorer discrimination of real vs. fake headlines under RH-bias stimulation compared to sham is driven by lower hit rates under RH-bias stimulation (**Figure 46**; RH>SH Wilcoxon signed-rank test, statistic=56.5, p=0.04). There was no difference in false alarm rates under RH-stimulation (RH>SH Wilcoxon signed-rank test, statistic=72.5, p=0.22).

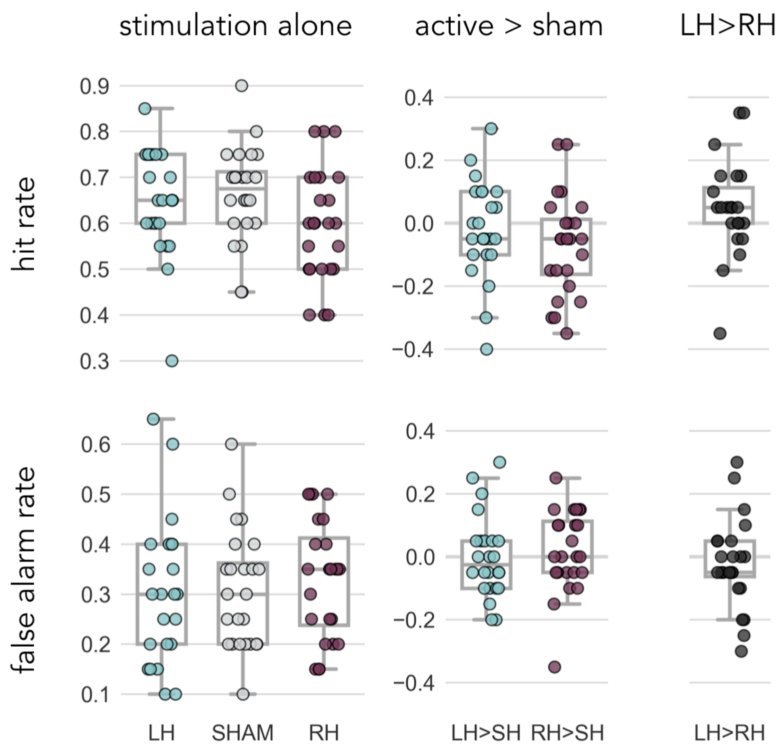

Figure 46: Condition differences in hit rates and false alarm rates.

Multiple linear regression was used to examine the influence of stimulation on plausibility ratings while accounting for additional factors (**Figure 47**). Contrary to predictions, RH-bias stimulation intensity was associated with greater plausibility ratings (beta: 0.09 ± 0.04, t=2.32, p=0.02), but the relationship was not significant after correcting for multiple comparisons. Consistency between the bias of the headline and participants' political leanings was associated with greater plausibility ratings. Finally, we found that people who reported trying harder rated headlines as less plausible and participants who were more strongly right-handed rated the headlines as more plausible.

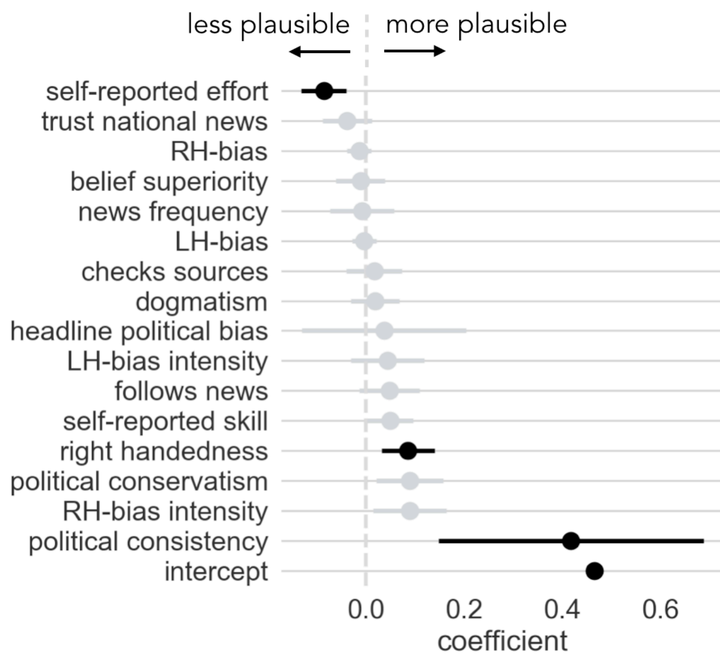

Figure 47: Linear regression results for headline plausibility. Values in black are significant after correcting for 17 comparisons with a Bonferroni correction ($t>2.98$, $p<0.003$).

We applied the same multiple regression model to participants' certainty ratings. In line with our framework, we expected LH-bias stimulation to be associated with more certainty and RH-bias stimulation to be associated with less certainty. We found that RH-bias stimulation intensity was associated with less certain plausibility ratings (beta: -0.05 ± 0.02, $t=-2.68$, $p=0.07$), but the comparison did not survive the multiple comparisons correction. Only participant-related factors influenced the certainty associated with participants' plausibility ratings: greater dogmatism and greater trust in national news organizations were associated with less certain ratings while right-handedness, self-reported effort, greater frequency of news consumption, and greater frequency of news source-checking were associated with greater certainty ratings.

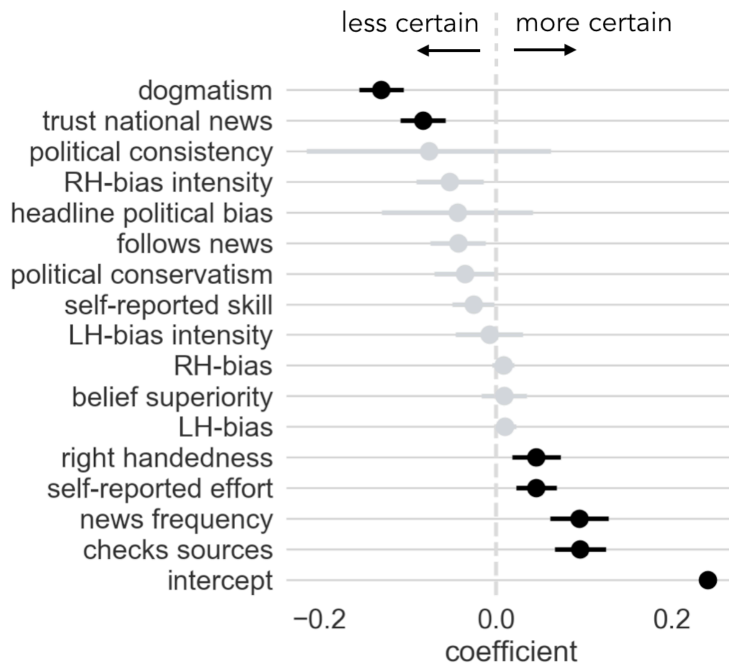

Figure 48: Linear regression results for headline rating certainty. Values in black are significant after correcting for 17 comparisons with a Bonferroni correction (t>2.98, p<0.003).

D. Discussion

The primary goal of this experiment was to determine how tDCS affected participants' belief updating under belief-evidence conflicts. In order to create more salient belief-evidence conflicts, we used scenarios that have more real-world applicability and that are naturally associated with stronger prior opinions. The ballot measures task in particular was designed to take advantage of the participants' pre-formed beliefs. Consistent with our predictions, the heatmap analyses provided evidence that RH-bias stimulation was associated with greater belief backtracks when evidence disconfirmed strong beliefs compared to sham and LH-bias stimulation in both the court case and ballot measure tasks (Figure 34 and Figure 41). A likely reason for observing greater belief-evidence conflict backtracks under RH-bias stimulation in this experiment but not the first tDCS experiment is that the state guessing task in the first experiment did not induce salient conflicts when disconfirmatory evidence was observed, whereas these tasks did.

In the multinomial regression analysis for the ballot measures, we found that participants were less likely to adopt more extreme beliefs when faced with conflicting arguments in the RH-bias stimulation condition. Previous research has shown that people with strong beliefs tend to paradoxically strengthen their belief when presented with evidence that contradicts it, which is likely due to discounting evidence that disconfirms strong beliefs and favoring evidence that confirms them (McKenzie, Lee, & Chen, 2002; Taber, Cann, Kucsova, 2009; Taber & Lodge, 2006, Redlawsk, 2002). The results from this experiment suggest that RH-bias stimulation abolishes, or at least mitigates, this tendency, since participants under RH-bias stimulation were less likely to advance their beliefs when they observed conflicting arguments in the ballot measure task.

We also found that RH-bias stimulation was associated with poorer discrimination between the real and fake headlines compared to sham stimulation, which was driven by lower hit rates. This suggests that RH-bias stimulation made participants less willing to rate headlines as being plausible, but did not change their evaluation of less plausible headlines. It is possible that these results can be explained by participants making less certain judgments under RH-bias stimulation, a trend that is reflected in the regression analysis but that did not reach significance after controlling for multiple comparisons. Since the average headline rating was slightly implausible (0.45 on a scale from 0 to 1), slight reductions in certainty would make slightly plausible headlines slightly implausible and implausible headlines slightly less implausible. Since we binarized participants judgments by considering ratings > 0.5 as "real" judgments and ratings < 0.5 as "fake" judgments, this would result in fewer "real" judgments for plausible headlines but would not affect the proportion of "fake" judgments for implausible headlines, thus resulting in a lower hit rate but normal false alarm rate.

Some differences in behavior were explained by individual difference measures. Most notably, we found that political conservatism was associated with greater backtracking and fewer non-updates in both the court case and ballot measure tasks. Since all 24 participants in this experiment leaned liberal, these results can be interpreted as participants with more moderate or mixed political views are more likely to backtrack on their beliefs than those with more extreme or more consistent liberal leanings. In a similar vein, we found that higher ratings of dogmatism were associated with more advances in the court case task.

The results of this experiment are generally in line with our predictions that RH-bias should be associated with greater backtracking when evidence conflicts strong beliefs and associated with establishing a higher threshold for accepting events to be real.

IV. Experiment 3: Hemispheric asymmetry during uncertainty and conflict

A. Rationale

As discussed previously, neuroimaging studies on healthy participants provide less support for hemispheric asymmetry in reasoning than do patient studies. Relying on existing neuroimaging studies to test our predictions of hemispheric asymmetry has several drawbacks. First, existing neuroimaging studies rarely, if ever, directly examine hemispheric asymmetry by comparing activity in the right and left hemispheres. Instead, one must use thresholded statistical maps to make inferences about hemispheric asymmetry, and doing so can lead to misleading conclusions. For example, a contrast in which values are just above the threshold in one hemisphere and just subthreshold in the contralateral hemisphere may make a process appear lateralized, even in the absence of significant differences in activity. In other cases, there may be activation clusters in both hemispheres, but the magnitudes of activity differ significantly. Second, most studies do not use contrasts that attempt to isolate the processes that we make predictions about, namely contrasts involving reducing uncertainty and reducing conflict. This limits our ability to use existing studies to test the predictions of our framework. Third, asymmetric processing may not produce asymmetric BOLD activity. Different neural network architectures may lead to different processing characteristics which in turn lead to different behavioral outcomes, but these functional asymmetries may go undetected with traditional neuroimaging approaches that examine BOLD activity profiles.

To address the first problem, we previously conducted a series of meta-analyses that directly compared right vs. left hemisphere activity across groups of contrasts in the existing reasoning literature. We grouped contrasts that were similar and that we expected to be more left- or right-lateralized. Much like the reasoning neuroimaging literature on the whole, the results were mixed. Some contrasts produced lateralized results in accordance with our predictions, but others produced bilateral results. Although our meta-analyses attempted to overcome the first problem by directly comparing left vs. right hemisphere activity, it was ill-suited to overcome the second problem because the contrasts that made up each meta-analysis were not designed to isolate the processes that we made predictions about. Due to the types of contrasts that fed into each meta-analysis, the results may have been driven more by ancillary processing demands (e.g., trial difficulty, visual vs. verbal processing, or working memory load) than the reasoning processes that we care about.

In order to better test our predictions about hemispheric asymmetry in reasoning, we ran an fMRI study that was designed to overcome both the first and second problems. Participants in the experiment completed the same state-guessing task described in Chapter 2. In the task, participants make inferences as evidence is presented sequentially. Due to the probabilistic nature of the task, participants' inferences are uncertain and their certainty levels change throughout each trial. Additionally, participants occasionally receive evidence that conflicts their current guess. This task allowed us to identify brain regions that are recruited when uncertainty is high or when conflict arises, which we predicted to be associated with greater left- and right-hemisphere activity, respectively. Furthermore, we directly tested our predictions by comparing activity in homologous regions in the left and right hemisphere for contrasts that we predict to have asymmetric activity.

Like other fMRI studies, this experiment does not address the third problem since differences in cognitive processing may not be associated with differences in BOLD activity. However, establishing asymmetry in contrasts that we predict to be asymmetric provides some evidence that neural processing in the left and right hemispheres is not identical and that the hemispheres play different roles in reasoning.

B. Methods

1. Participants

37 individuals (25 females, age mean: 21 years, age range: 18 to 37 years) participated in the functional magnetic resonance imaging (fMRI) experiment. Participants were recruited from an online participant pool, through recruitment emails sent to undergraduate mailing lists, and through word of mouth. All participants had to pass a behavioral prescreen before being scanned. In the prescreen, participants filled out eligibility forms for fMRI, TMS, and tDCS and completed two behavioral tasks: a probabilistic inference task (described in Chapter 2) and a recognition memory task for a separate experiment. Participants who failed to shift their decision criteria in the memory task, failed to discriminate studied vs. unstudied items in the memory test above chance, or failed to move the slider in at least 50% of the first 6 observations in the probabilistic inference task were excluded from participating in the fMRI experiment. Participants who were deemed unsafe or ineligible to participate in the fMRI, TMS, or tDCS experiments were also excluded. Eligibility criteria included 1) no metal in the body other than dental work, 2) no history of epilepsy, stroke, or brain damage, 3) no pacemaker or brain stimulator, 4) no dredlocks or irritation on the scalp near the electrode sites, and 5) not currently pregnant. Participants were paid $20 per hour for their participation in the experiment.

One participant withdrew from the fMRI experiment before the functional scans were collected due to feelings of claustrophobia. The dataset consisted of data from the remaining 36 participants.

2. Task

The same probabilistic inference task described in Chapter 2 was used in the fMRI experiment. Briefly, participants were told that the computer randomly selected one of two U.S. states and their goal was to guess which state was selected based on the ethnicities of people who live in the selected state. Participants observed residents' ethnicities one at a time for 2.8 seconds (7 TRs) each and reported their current guess about which state was selected with a continuous digital slider. Participants used the left and right buttons of an MRI-compatible button box to move the slider leftward and rightward. Unlike the tDCS experiment, participants were not given the option to stop collecting evidence in the fMRI experiment. Ten residents were presented in every trial and participants were required to view all ten. Participants also did not report their final guess or their confidence at the end of each trial. To encourage participants to use the slider thoughtfully, bonuses were awarded based on the position of the slider (described in more detail in Chapter 2).

3. Procedure

At the beginning of every scanning session, participants signed the consent form, read task instructions, completed a practice session, and changed into hospital scrubs. After verifying that participants removed all metal, they were assisted into the scanner. Participants were instructed to remain as still as possible during each scan and were given neck pads to help keep their heads still.

Each fMRI session consisted of 15 scans, which were collected in the following order: 1 localizer scan, 1 gradient echo field map, 1 t2-weighted anatomical scan, 5 functional localizer scans (300 TRs), 3 functional scans for the state guessing task (1060 TRs each), 1 t1-weighted anatomical scan, and 2 functional scans for a memory task. Each scanning session took 1.25-1.5 hours.

The state guessing functional scans resembled a slow-event related design. Each of the ten trials per functional run were separated by 12 seconds (30 TRs). Within each trial, the 10 ethnicities were presented continuously for 2.8 seconds each, for a total of 28 seconds. The scans were set up this way to allow for analysis of the temporal dynamics of each trial in isolation.

4. Data Acquisition

MRI data was acquired at the Brain Imaging Center at the University of California, Santa Barbara on a Siemens 3T Prisma scanner with a 64-channel phased-array head coil. Blood Oxygenated Level Dependent (BOLD) contrasts were measured with a gradient-echo echoplanar imaging sequence (400ms repetition time (TR); 35ms echo time (TE); 52 degree flip angle (FA)). Each volume consisted of 48 slices acquired parallel to the AC-PC plane that covered the whole brain (multi-band acceleration factor of 8, interleaved acquisition; 3 mm slice thickness; 3 x 3 mm in-plane resolution; 192mm field of view (FOV); 64 x 64 matrix). The T1-weighted anatomical scan was collected sagittally with an anterior-to-posterior phase encoding direction (0.94mm slice thickness, MPRAGE, 2500ms TR; 2.22ms TE; 7 degree FA; 241 mm FOV). The T2-weighted anatomical scan was also collected sagittally with an anterior-to-posterior phase encoding direction (0.94mm slice thickness, MPRAGE, 3200ms TR; 566ms TE; 241 mm FOV).

5. Data Processing

Preprocessing and statistical analyses were carried out with the fMRI Expert Analysis Tool (FEAT, version 5.0.10), which is part of the FMRIB software Library (FSL, http://fsl.fmrib.ox.ac.uk/fsl). The data in each functional run were filtered with 100s high-pass filter, motion corrected with MCFLIRT, and spatially smoothed with a 5mm full-width at half-maximum kernel. BET (Brain Extraction Tool) was used to remove non-brain voxels, field map unwarping was applied to account for inhomogeneity in the magnetic field, and FILM pre-whitening was used to reduce the autocorrelation in the data. The functional images were registered to brain-extracted T1 anatomical scans using the linear boundary-based registration (BBR) algorithm. The anatomical images were registered to the standard MNI-152 2mm brain template using an affine transformation with 12 degrees of freedom using FLIRT, followed by a nonlinear transformation with a 10mm warp resolution using FNIRT.

6. Univariate Data Analysis

The design matrix for the general linear model consisted of regressors for belief certainty, belief-evidence conflict, belief updates, and evidence expectedness, as well as boxcar regressors for the different TR types (**Table 1**). Evidence strength was operationalized as the absolute value of the log likelihood ratio:

$$evidencestrength = \left| log\left(\frac{P(ethnicity|state_{guessed})}{P(ethnicity|state_{alternate})}\right)\right|$$

Equation 3: Estimate of evidence strength. Higher values are indicative of stronger evidence. The likelihood estimates were taken from participants' estimates of state demographics. The "guessed state" was defined as the state that the slider was closest to in the previous TR. Log likelihood ratios greater than 0 were considered confirmatory evidence and log likelihood ratios less than 0 were considered disconfirmatory evidence.

Evidence expectedness captured how expected the ethnicity evidence was based on participants' state hypotheses and demographic estimates:

$$P(eth) = P(eth|state_{guessed}) * P(state_{guessed}) + P(eth|state_{alt}) * P(state_{alt})$$

Equation 4: Estimate of evidence expectedness.

Table 1: Regressors in fMRI analysis.

Regressor	Type	Demean	Description
prior certainty	continuous	yes	certainty at previous TR
belief update: advance	continuous	no	magnitude of update toward slider ends
belief update: backtrack	continuous	no	magnitude of update toward slider middle
disconfirmatory evidence (DE)	continuous	no	strength of evidence (absLLR) that conflicted prior belief

confirmatory evidence (CE)	continuous	no	strength of evidence (absLLR) that was consistent with prior belief
DE * prior certainty	continuous	no	interaction of conflicting evidence and prior certainty regressors
CE * prior certainty	continuous	no	interaction of supporting evidence and prior certainty regressors
evidence expectedness	continuous	yes	bayesian estimate of expectedness of ethnicity: P(ethnicity)
button press	boxcar	no	presence of button press
resident presentation	boxcar	no	presentation of resident
introduction presentation	boxcar	no	presentation of state options
new evidence	boxcar	no	TRs when new resident was presented

absLLR: absolute log likelihood ratio, abs(log(P(ethnicity|state) / P(ethnicity|~state)))

Temporal derivatives and six standard motion parameters were added to the model for a total of 30 regressors. Regressors were convolved with a Gamma hemodynamic response function (HRF) with 0s phase, 3s standard deviation, and 6s lag.

A first-level analysis was carried out on every participant and every run. Within each participant, parameter estimates were averaged across runs with a standard-weighted fixed-effects model. Participants' parameter estimates were averaged to produce group-level maps for every contrast using a mixed effects model (FLAME 1 in FEAT). At every level, multiple comparisons were corrected for by applying an initial z-threshold of 2.3 and a cluster threshold of 0.05.

7. Hemispheric asymmetry analysis

A hemispheric asymmetry analysis was conducted to directly test for functional asymmetries between homologous regions in the left and right frontal lobes for contrasts that were predicted to produce asymmetric activations. Anatomical regions consisted of all of the prefrontal structures in the Harvard-Oxford cortical atlas and included the frontal pole (FP), superior frontal gyrus (SFG), middle frontal gyrus (MFG), inferior frontal gyrus pars triangularis (IFGpt), and inferior frontal gyrus pars opercularis (IFGpo). Mid-level contrast of parameter estimates (cope) maps were masked with each of the anatomical ROIs and the average value within each mask was computed for the left and right hemispheres. A repeated-measures t-test was used to test for differences between left vs. right hemisphere activity for each contrast and each anatomical region. The analysis consisted of 45 comparisons (5 anatomical regions x 9 contrasts). A Bonferroni correction was applied to account for multiple comparisons.

C. Results

1. Contrasts with predicted left-lateralized frontal activity

As before, certainty was operationalized as the distance away from the middle of the slider track, and ranged from 0 (completely uncertain) to 0.5 (completely certain). We predicted that neural activity in left frontal areas would be commensurate with uncertainty at the beginning of each TR, such that greater uncertainty would be associated with greater left frontal activity. Many brain regions were more active with increasing uncertainty, including the left middle frontal gyrus (MFG), medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), occipital cortex (OCC), left motor and sensory cortices, bilateral insula, left posterior parietal cortex (PPC), and basal ganglia. Only the right sensory cortex was associated with greater activity as certainty increased (**Figure 49**).

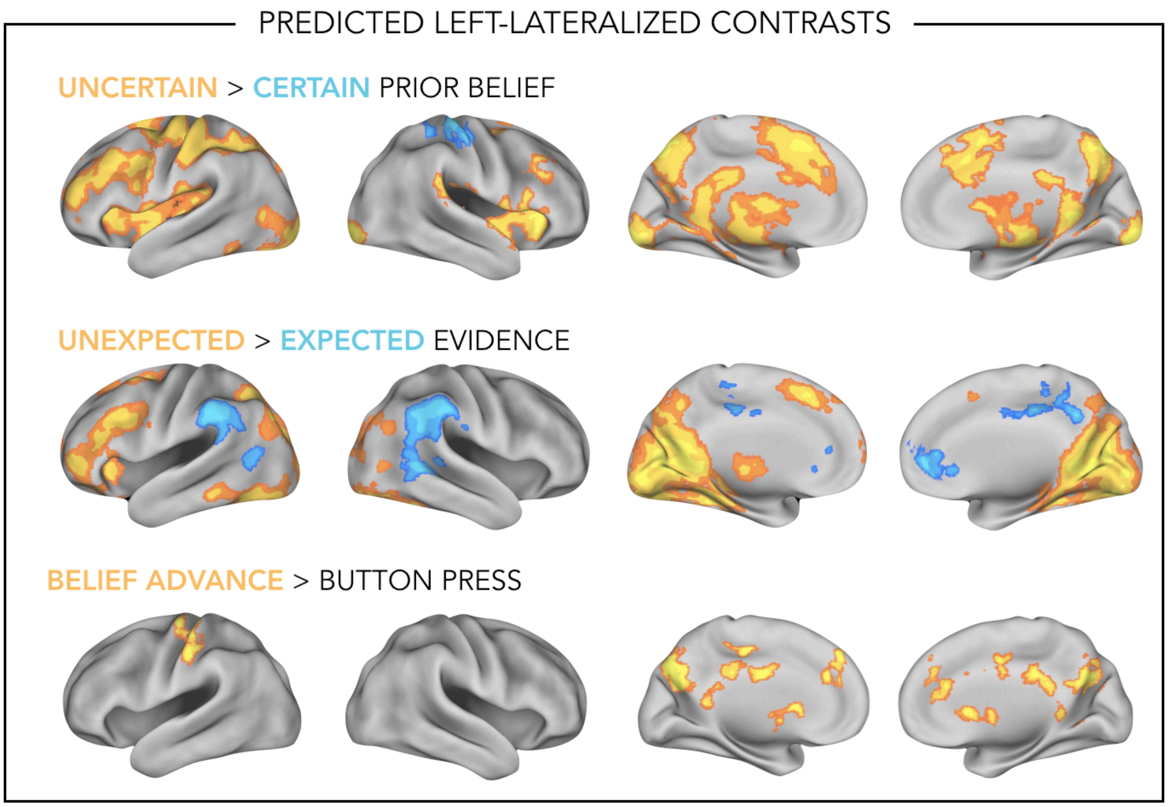

Figure 49: Contrasts with predicted left-lateralized frontal activity. Maps are thresholded at z=2.3.

A second contrast examined relationships between brain activity and evidence expectedness. For each observation, the expectedness of the resident's ethnicity, P(ethnicity), was calculated using participants' demographic estimates and their prior beliefs that each state was selected (as indicated by their slider position at the previous TR). We predicted that unexpected ethnicities would be associated with left frontal activity, since they would introduce uncertainty. We found that unexpected ethnicities were associated with greater activity in the left superior, middle, and inferior frontal gyri, left insula, left mPFC, and bilateral occipital cortex, while expected ethnicities were associated with greater activity in the bilateral supramarginal and angular gyri.

Finally, since we propose that left-lateralized networks are driven toward reducing uncertainty, we predicted that belief advances (in which the slider is moved toward either end of the track) would be associated with left-lateralized activations because they reduce uncertainty. We found that, compared to all button presses, belief advances were only associated with greater activity in left motor and sensory regions, the posterior cingulate cortex, dorsomedial cortex, and right caudate.

2. Contrasts with predicted right-lateralized frontal activity

Contrasts the captured conflicts between beliefs and evidence were predicted to have greater right vs. left frontal activity. The first contrast identified brain areas that were preferentially active during the presentation of confirmatory or disconfirmatory evidence. Confirmatory evidence consisted of ethnicities that were more prevalent in the guessed state (closer to the slider) than the rejected state (farther from the slider) based on the participants' estimates of state demographics. Similarly, disconfirmatory evidence consisted of ethnicities that were more prevalent in the rejected state than the guessed state. Evidence strength increased as the discrepancy between ethnicity estimates in the guessed and rejected states increased. No brain regions became significantly more active as the strength of disconfirmatory evidence increased nor as the strength of confirmatory evidence increased.

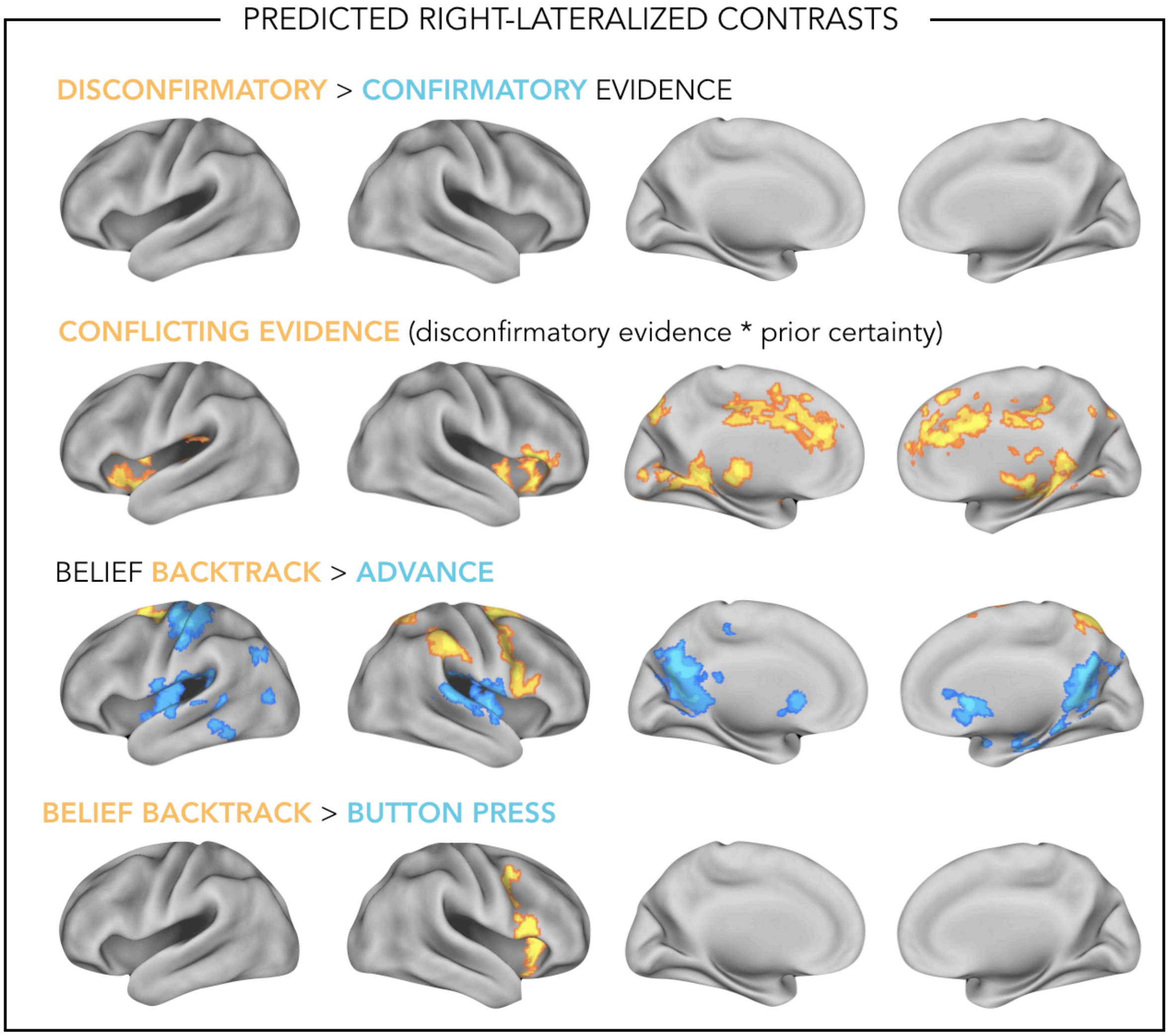

Figure 50: Contrasts with predicted right-lateralized frontal activity. Maps are thresholded at z=2.3.

One tenet of the hemispheric asymmetry framework is that right frontal neural networks are recruited when evidence disconfirms strong prior beliefs. To identify brain regions that are recruited when evidence conflicts strong beliefs, a second contrast was carried out that examined the interaction between disconfirmatory evidence and prior belief certainty. The bilateral insula, bilateral anterior cingulate cortices (ACC), right IFG, bilateral lingual gyri, and bilateral thalami were more active as belief-evidence conflict increased. Contrasting strong belief-evidence conflicts (disconfirmatory evidence x prior certainty) vs. strong belief-evidence argreements (confirmatory evidence x prior certainty) resulted in greater activations in the bilateral ACC and right thalamus.

Finally, we proposed that right frontal networks may play a funtional role in reversing beliefs that are no longer consistent with the evidence. To test this hypothesis, all slider movements were categorized as either an advance, in which the slider was moved closer to either end of the track, or a backtrack, in which the slider was moved closer to the middle of the track. A contrast was carried out that compared the magnitude of backtracks vs. the magnitude of advances. The left motor cortex, bilateral lingual gyri, and bilateral insula were more active with greater advances than greater backtracks. The right IFG, bilateral SFG, right supramarginal gryus, and right posterior parietal cortex (PPC) were more active with greater backtracks than greater advances. Similarly the right anterior insula, right IFG, and right MFG are more active during belief backtracks compared to button presses.

3. Hemispheric asymmetry analysis

In order to directly test which hemisphere had greater activations in contrasts that we expected to be more right-lateralized or left-lateralized, we ran multiple repeated-measures t-tests that compared activations in left and right frontal anatomical regions for four contrasts with predicted left-lateralized activity and five contrasts with predicted right-lateralized activity. The five frontal ROIs were extracted from the Harvard-Oxford atlas and included the frontal pole (FP), superior frontal gryus (SFG), middle frontal gyrus (MFG), inferior frontal gyrus pars triangularis (IFGpt), and inferior frontal gyrus pars opercularis (IFGpo). We expected that contrasts for uncertainty, unexpected evidence, and belief advances would be associated with greater left frontal activity, and found that 15 out of 20 comparisons were associated with greater left vs. right activity. Both of the inferior frontal gyri ROIs were associated with greater right hemisphere activity for the uncertainty contrast and advance > button press contrast. All frontal ROIs were associated with significantly greater activity in the left vs. right homologues for the unexpected evidence contrast after applying a Bonferroni correction to correct for multiple comparisons.

	Anatomical Region				
	FP	SFG	MFG	IFGpt	IFGpo
Predicted left-lateralized contrasts					
uncertainty	2.51	2.63	1.02	-1.53	-0.40
unexpected evidence	**3.59**	**5.92**	**4.36**	**4.26**	**4.46**
advance > backtrack *	2.09	2.69	1.96	**3.79**	2.99
advance > button press	0.38	2.17	-0.54	-0.36	-1.46
Predicted right-lateralized contrasts					
disconfirmatory > confirmatory evidence	1.46	0.15	0.86	1.59	0.52
disconEv*certainty	-0.80	0.61	0.26	-1.38	0.15
disconEV*cert > conEv*cert	-1.63	-0.74	-0.68	-1.67	-0.64
backtrack > advance *	-2.09	-2.69	-1.96	**-3.79**	-2.99
backtrack > button press	-1.37	-1.52	-2.76	**-3.88**	-3.34

Figure 51: T-values from repeated-measures t-tests comparing average activation in left vs. right anatomical regions for each contrast. Anatomical regions include the frontal pole (FP), superior frontal gyrus (SFG), middle frontal gyrus (MFG), inferior frontal gyrus pars triangularis (IFGpt), and inferior frontal gyrus pars opercularis (IFGpo). Negative values (purple) indicate more right-lateralized activity and positive values (blue) indicate left-lateralized activity. Values in bold are statistically significant after accounting for 45 comparisons (p<0.0011). Note: the advance > backtrack contrast and the backtrack > advance contrast are opposite contrasts, so the lateralization values are symmetric.

Five contrasts, all of which involved disconfirmatory evidence or belief backtracks, were predicted to be associated with more right-lateralized activity. Of the 25 comparisons, 17 were associated with greater activations in the right hemisphere. Only the inferior frontal gyrus pars triangularis was associated with significant right-lateralized activity for the two backtrack contrasts after correcting for multiple comparisons. Opposite of predictions, the disconfirmatory > confirmatory evidence contrast was associated with greater activity in the left vs. right hemisphere in each anatomical region, although none of the comparisons reached significance.

27 out of the 40 predictions were in the predicted direction, which is significantly above chance (binomial test: $z=2.05$, $p=0.02$; note: the pair of advance/backtrack contrasts were counted as one set of predictions since it is the same contrast). If the disconfirmatory > confirmatory evidence contrast is excluded (since it involves disconfirmatory evidence irrespective of prior beliefs and our predictions specifically regard disconfirmatory evidence and strong prior beliefs), 27 of 35 comparisons are in the predicted direction (binomial test, $z=3.04$, $p<0.001$).

E. Discussion

1. Summary of results

The goal of this study was to use fMRI to test our predictions that the left hemisphere plays a dominant role in reducing uncertainty and the right hemisphere plays a dominant role in resolving conflict. We compared activity in left vs. right frontal anatomical regions for contrasts that we expected to be more left- or right-lateralized. On the whole, the fMRI results were consistent with our predictions. We found that the contrasts that involved heightened uncertainty or the reduction of uncertainty were largely associated with greater activity in left frontal areas compared to right frontal areas. Only the inferior frontal gyrus was associated with greater right hemisphere activity for the uncertainty and advance > button press contrasts. Likewise, we found that the contrasts that involved conflict or belief backtracking were associated with greater activity in the right vs. left hemisphere in most frontal anatomical regions. Not all of the differences were significant after correcting for multiple comparisons, but all of the comparisons that were significant were consistent with our predictions.

2. Conflicting vs. disconfirmatory evidence

The fMRI results highlight the difference between processing disconfirmatory evidence and conflicting evidence. Here, we define disconfirmatory evidence as all evidence that provides more support for the alternative hypothesis than the current hypothesis. When participants only have a weak hunch about which state was selected, disconfirmatory evidence is informative and can be used to reduce uncertainty. However, when prior beliefs are strong and participants feel certain they know which state was selected, disconfirmatory evidence produces conflict and may require backtracking from the current belief. We refer to evidence that disconfirms strong prior beliefs as conflicting evidence. Greatest conflict occurs when both the prior belief is strong and the evidence strongly supports the alternative hypothesis, and conflict decreases as either the prior belief strength decreases or the evidence strength decreases. We found right-lateralized activations for conflicting evidence, but left-lateralized activations for disconfirmatory evidence. This highlights the role of right frontal regions in processing belief-evidence conflicts. These results support the idea that the left hemisphere plays a dominant role in inference making; when uncertainty is high and disconfirmatory evidence aids hypothesis selection, left-frontal networks dominate. Right frontal networks are only recruited when evidence disconfirms a strongly held belief, supporting the idea that right frontal networks figuratively apply a "cognitive brake" when beliefs and evidence conflict.

3. Belief advances vs. belief backtracks

A central tenet of our framework is that left frontal networks are biased toward reducing uncertainty and right frontal networks are biased toward reducing conflict. We argue that these biases contribute to asymmetric processing in reasoning, such that left frontal networks are recruited to reduce uncertainty (for example, during inference making or hypothesizing) and right frontal networks are recruited to reduce conflict (for example, when evidence conflicts strongly held beliefs). In this fMRI experiment, we recorded participants' brain activity and tracked when and to what extent their beliefs changed as they observed evidence. According to our framework, we should expect to find more left-lateralized activity when participants become more certain of their guesses and move the slider closer toward either end of the track. In contrast, we should expect to find more right-lateralized activity when people backtrack on their guess and move the slider closer toward the middle of the track. Consistent with our predictions, the advance > backtrack contrast was associated with greater activity in the left vs. the right homologues of each of the five frontal anatomical regions we examined. This provides support for our proposal that left frontal networks play a dominant role in reducing uncertainty by making inferences and right frontal networks play a dominant role in reducing conflict by backtracking on untenable hypotheses.

Since both belief backtracks and belief advances were types of slider updates and participants could only update the slider by pressing buttons, we attempted to subtract out motor-related activity due to button presses. Even so, we found motor-related activity in the advance > button press contrast. This is likely due to the fact that the regressor for belief advances was continuous and increased with larger advances. The regressor for button-presses, on the other hand, was a boxcar (in order to avoid creating a singular design matrix). Since the belief advance regressor captured both the presence and durations of button presses while the button press regressor only captured the presence of button presses, it is likely that the button press regressor did not fully account for all motor related activity, which may explain the motor-related activity in the advance > button-press contrast.

4. Uncertainty and the inferior frontal gyrus

Surprisingly, the inferior frontal gyrus was associated with relatively greater activity in the right hemisphere than the left hemisphere for the uncertainty contrast. Not only is this result inconsistent with our predictions, but it also complicates the interpretability of the tDCS experiments. We targeted the bilateral inferior frontal gyri under the assumption that the left IFG was more sensitive to uncertainty and the right IFG was more sensitive to conflict. Although we did find that activity increased in the left IFG with increasing uncertainty, it did so to a lesser extent than it did in the right IFG. This may explain why some behavior was occasionally more similar between the LH-bias and RH-bias conditions in the tDCS experiment than between each active condition and the sham condition.

Based on the results of this experiment, it may be beneficial to repeat the tDCS experiments with a modified electrode montage. One option is to introduce more electrodes in order to stimulate the left and right frontal lobes more diffusely. Another option is to target the right IFG as before, but change the left hemisphere target to the middle frontal gyrus since it was robustly activated during uncertainty. These issues are discussed in more detail in the next section.

V. Discussion

A. Summary of the results

Overall, we found more evidence for our predictions regarding the effects of RH-bias stimulation than LH-bias stimulation. In the first tDCS experiment, we found that participants who received more intense RH-bias stimulation at the cortex collected more evidence, adopted higher thresholds for stopping evidence collection, and were less certain than ideal throughout the beginning of the evidence presentation. In the second tDCS experiment, we found that RH-bias stimulation was associated with a smaller likelihood of belief advances (and corresponding greater likelihoods of backtracks and non-updates) during conflict in the ballot measures task. The heatmap analyses for the criminal court cases and ballot measures revealed that RH-bias stimulation was associated with greater backtracking when evidence disconfirmed certain beliefs compared to sham stimulation and, to a lesser extent, compared to LH-bias stimulation. Under RH-bias stimulation, participants were less likely to judge real news headlines as being real, which resulted in poorer discrimination of real vs. fake headlines compared to sham and LH-bias stimulation. Finally, in the fMRI experiment, we found that right frontal regions were recruited more robustly than left frontal regions during conflict and during belief backtracks.

Greater LH-bias stimulation intensity was only associated with one behavioral measure. We found that greater LH-bias stimulation intensity led to more evidence collection in the state guessing experiment, which was inconsistent with our predictions. Surprisingly, we consistently found that the influence of LH-bias stimulation intensity on behavior was in the same direction as the influence of RH-bias stimulation intensity, but was always associated with a smaller effect size. The results from the fMRI experiment regarding left-lateralized contrasts were generally in line with predictions. We found that the left frontal anatomical regions were more strongly recruited than right frontal regions for contrasts that we expected to be more left-lateralized, including greater uncertainty, advances in beliefs, and unexpected evidence. However, we found that the inferior frontal gyrus, which we stimulated with tDCS, was associated with greater activity for the uncertainty and advance > button press contrasts, which is inconsistent with expectations.

B. Limitations, interpretations, and implications

1. Electric field strengths at the cortical surface

Voroslakos and colleagues (2018) recently found that electric fields of 1 V/m or greater are required to modulate neuronal firing in rats. This finding challenges the widely accepted view that tDCS modulates spontaneous neural activity because models of current flow indicate that common tDCS electrode montages induce electric fields weaker than 1 V/m at the cortical surface.

Since we modeled the current flow for every participant in both tDCS experiments, we can estimate the strength of the electric fields induced at the cortical surface and determine how many participants had cortical electric fields greater than 1 V/m (**Figure 52**). In the first tDCS experiment, we induced electric fields greater than 1 V/m at the cortical surface for 17 out of the 26 participants (65.4%). In the second tDCS experiment, we only induced cortical electric fields greater than 1 V/m for 3 of the 24 participants (12.5%). The large disparity between the two tDCS experiments is the result of using 2mA stimulation in the first tDCS experiment and 1.5mA stimulation in the second. Stimulation intensity was purposely reduced in the second experiment as a safety precaution. Since the stimulation duration in the second experiment was almost twice as long as the first experiment, maintaining a current intensity of 2mA would nearly double the total amount of current delivered. Second, the current modeling results from the first tDCS experiment indicated that the cortical current densities induced by our electrode montage were higher than the current densities induced by common HD-montages and sponge montages (Bikson et al, 2016), so we felt it was prudent to reduce the stimulation intensity so that the cortical current density was more in line with the literature (Note: this decision was made prior to the publication of Voroslakos et al. (2018)).

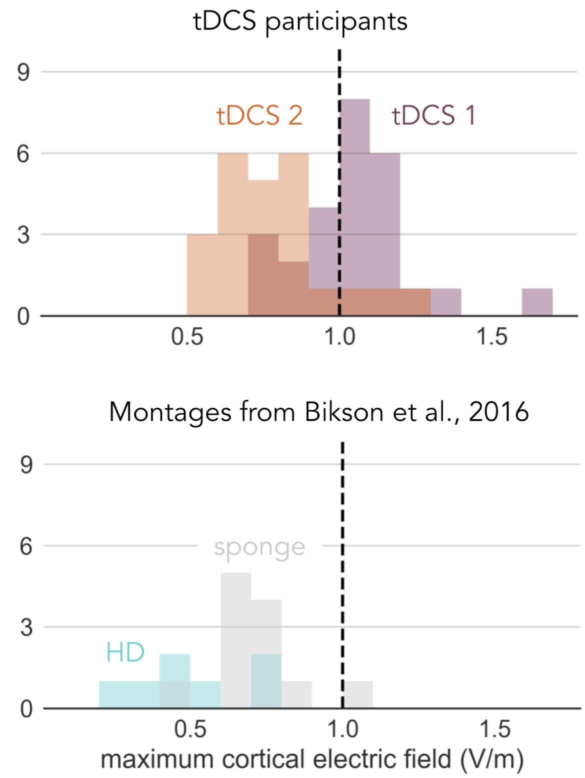

Figure 52: Maximum electric field magnitudes at the cortical surface (V/m). Top: Estimated electric field magnitudes for every participant in the first (purple) and second (orange) tDCS experiments. Bottom: Estimated electric field magnitudes for the HD-tDCS (blue) and sponge (gray) montages reported in Bikson et al., 2016.

In their 2016 review of the safety and tolerability of tDCS, Bikson and colleagues present the maximum current density on the skin and brain surfaces for 7 HD montages and 12 traditional sponge montages. In order to compare the electric field intensities induced by these montages to those induced by our electrode montages, we determined the slope and intercept needed to linearly map current density to electric fields based on our modeling data (linear regression with a slope and intercept term, R^2=1), estimated the cortical current densities from the scatter plot in Figure 3 of Bikson et al. (2016), and converted the current densities to electric fields. **Figure 52** includes the distribution of cortical electric fields for the HD-montages (blue) and sponge montages (gray). Only one montage is associated with an electric field intensity at the cortical surface greater than 1 V/m and that montage was modeled with a pediatric head.

The electric fields we induced in the first tDCS experiment are greater than those induced by common electrode montages, and the electric fields we induced in the second tDCS experiment are well within the range of those induced by traditional sponge montages. The reason that our electrode montage was associated with stronger electric fields than the other HD montages is that our montage consisted of electrodes on each side of the head, which served to pull current through the brain and maximize current intensity at the cortical targets. Most of the other HD montages used a 4x1 ring configuration (with an anode surrounded by four cathodes) which is intended to maximize current focality rather than intensity. Even though our current modeling results indicate that we did not induce electric fields greater than 1 V/m at the cortex for all participants (especially in the second tDCS experiment), the electric fields that we induced were as great or greater than the fields induced by electrode montages that are commonly reported in the literature. Still, the differences in cortical stimulation intensities between experiments 1 and 2 may explain the stronger effects of stimulation on behavior in the first experiment compared to the second experiment.

2. Sources of variability in tDCS and the need for modeling

In the last two years, several studies have been published that paint more nuanced pictures of the mechanisms and effects of tDCS. Although it has been known for a long time that the effects of tDCS depend on electrode placement, electrode size and shape, current intensity, and stimulation duration (Nitsche et al., 2008), more recent research has shown that the effects of tDCS depend on additional factors that are less controllable. Some of the sources of variability in tDCS studies, both controllable and uncontrollable, are listed in Table 2.

Kronberg, Bridi, Abel, Bikson, and Parra (2016) applied electric fields to hippocampal slices and found that tDCS only modulated synaptic plasticity of neurons that were already active. They also found that cathodal stimulation reduced long term depression (LTD) and enhanced long term potentiation (LTP) in apical dendrites, whereas anodal stimulation reduced LTD in apical dendrites but enhanced LTP in basal dendrites. The asymmetric effects of stimulation on synaptic plasticity complicates the simple notion that anodal stimulation is excitatory and cathodal stimulation is inhibitory. Rahman et al. (2013) examined how the orientation of neurons in an electric field influences the effects of tDCS. They found that radial electric fields, in which the neurons run parallel to the field, produce the canonical effects of excitatory anodal stimulation and inhibitory cathodal stimulation, but transverse fields that run perpendicular to the neurons create asymmetric effects that depend on neuronal pathways. Lafron, Rahman, Bikson, and Parra (2016) also found evidence that neurons increase firing under anodal stimulation and decrease firing under cathodal stimulation when aligned with the electric field, but they reported asymmetric effect sizes. Anodal stimulation was associated with much stronger effects than cathodal stimulation. One important caveat of these studies is that they used electric fields that are order of magnitudes stronger than those induced by tDCS in humans (8-35 V/m compared to < 1 V/m), so these mechanisms may not explain how tDCS exerts its effects in humans.

There's also no clear evidence that increasing stimulation intensity corresponds to greater changes in neuronal activity or in behavior. In a recent study, Jamil et al. (2017) characterized the after-effects of anodal and cathodal tDCS at different stimulation intensities. They found that anodal stimulation reliably increased cortical excitability, but cathodal stimulation only inhibited cortical excitability at 1mA, but not at 0.5mA, 1.5mA or 2mA. Furthermore, although anodal stimulation consistently increased cortical excitability, there was no dose-dependent relationship between current intensity and changes in cortical excitability. In a review of dosing effects in tDCS, Esmailpour et al. (2017) conclude that there is no reliable evidence showing that there is a linear, or even a monotonic, relationship between current intensity and neuronal effects.

In addition to stimulation parameters, neuronal orientation, and current intensity, the effects of tDCS can be influenced by individual differences in head size (Bikson et al., 2016; Kessler et al., 2013), fat content (Truong et al., 2013), skull thickness (Opitz, Paulus, Antunes, & Theilscher, 2015), cerebrospinal fluid volume and thickness (Lassko, Tanaka, Koyama, Santis, & Hirata, 2015; Opitz et al., 2015), sulcal depth (Opitz et al., 2015), age (Lassko et al, 2015; Fujiyama et al., 2014; Heise et al., 2014), and baseline cognitive abilities (Jones & Berryhill, 2012; Hsu, Juan, & Tseng, 2016; Rosen et al., 2016). In their review of the factors that contribute to individual variability tDCS responses, Li, Uehara, and Hanakawa (2015) provide evidence that the effects of tDCS are influenced by anatomy, functional organization of local neuronal circuits, baseline motor and cognitive abilities, baseline neurochemistry, circadian rhythms, and genetics. Additionally, several studies have shown that task demands can interact with the effects of stimulation (Bortoletto, Pellicciari, Rodella, Miniussi, 2015; Cabral et al., 2015; Gill, Shah-Basak, & Hamilton, 2015; Saucedo Marquez, Zhang, Swinnen, Meesen & Wenderoth, 2013).

Table 2 : Controllable and uncontrollable sources of variability in tDCS.

Sources of variability in tDCS
Controllable sources of variability
Electrode montage and stimulation parameters
Electrode size and shape
Configuration of anode(s) and cathode(s)
Current intensity
Stimulation duration
Stimulation procedures
Electrode placement between sessions and between participants
Uncontrollable sources of variability
Neuronal excitability
Endogenous neural activity
Orientation of neurons in the electric field
Neuron morphology and the distributions of different cell types
Age
Network effects
Non-linear dose responses to current intensity
Delivery of current
Volume and flow of cerebrospinal fluid
Tissue conductivity
Sulcal depth
Blood pressure
Resistance at electrode sites
Head size
Fat content
Skull thickness
Networks and behavior
Cognitive abilities at baseline
Cognitive strategies
Expectations
Task demands
Compensation

Current may also be distributed unintuitively throughout the brain. It has repeatedly been shown that sponge electrode montages deliver peak current densities to cortical regions in between the anode and cathode rather than beneath the anode as originally thought. It has also recently been shown that tDCS can deliver current to deep brain structures due to the high conductivity of cerebrospinal fluid. Huang et al. (2017) recorded electric fields with intracranial electrodes embedded in the brains of epileptic patients as they were stimulated with tDCS and found that electric fields in deep brain areas were nearly as intense as electric fields on the cortical surface.

These sources of variability muddle the simplistic view that anodal stimulation increases cortical activity and cathodal stimulation decreases it and point to the importance of using current modeling to estimate the effects of stimulation on an individual basis. We went to great lengths to control for some of these sources of variability. We defined neural targets for each participant based on their cortical folds, we modeled current flow in each participant to account for differences in head size, skull thickness, and fat content in the head, and we used a camera positioning system to ensure that the electrodes were placed precisely and consistently across sessions. Finally, we accounted for individual differences in maximum cortical current density in all of our regression analyses. We are unaware of any tDCS or HD-tDCS study that was as well controlled as ours.

We repeatedly found that individual differences in the intensity of the stimulation at the cortical surface predicted differences in behavior. This suggests that, not only is current delivered to the cortex differently in different people, but that these differences are important and can influence behavior. Our results provide strong evidence that researchers should model current flow for all participants and include the modeling results as factors in their analyses.

3. Similarity between RH-bias and LH-bias stimulation conditions

We predicted that LH-bias stimulation and RH-bias stimulation would produce opposite and symmetric effects, such that if LH-bias stimulation was associated with greater effects compared to sham, RH-bias would be associated with smaller effects (or null effects) compared to sham, and vice versa. However, we often found that behavior in the two active stimulation conditions were more similar than either condition compared to sham. For example, in each of the continuous heatmaps, the LH>sham and RH>sham maps generally looked similar – certainly more similar than would be expected if LH-bias stimulation and RH-bias stimulation induced opposite effects compared to sham. Furthermore, in almost all of the regression analyses, the coefficients for stimulation intensity went in the same direction for both the RH-bias and LH-bias stimulation conditions, which suggests that the magnitude of stimulation had a greater effect than the polarity of stimulation. It is important to note that in all but one case, the effects of LH-bias stimulation were not significant, even before correcting for multiple comparisons. Even though the effects were not significant, the fact that they consistently went in the same direction as RH-bias stimulation warrants a discussion.

There are several possible explanations for these results that can be summarized by three general explanations: 1) the physical sensations produced by active stimulation exert a stronger influence on behavior than any changes in neural activity, 2) our predictions about hemispheric asymmetry in reasoning are wrong, or 3) our assumptions about the neuromodulatory effects of tDCS are wrong. We discuss the merit of each of these explanations below.

The first interpretation of the similarity between LH-bias and RH-bias conditions is that the physical sensations of the stimulation are driving changes in behavior. It is possible that greater current densities produce greater physical sensations and these sensations are salient enough during the task to influence behavior. To test this hypothesis, we examined the relationship between the estimated cortical current density for each participant and their reports of the intensity of the tDCS sensations. We found no relationship between participants' sensation intensity ratings and their estimated cortical current densities (Figure 53), suggesting that the intensity of sensations associated with tDCS is not driving the differences observed between the active and sham conditions.

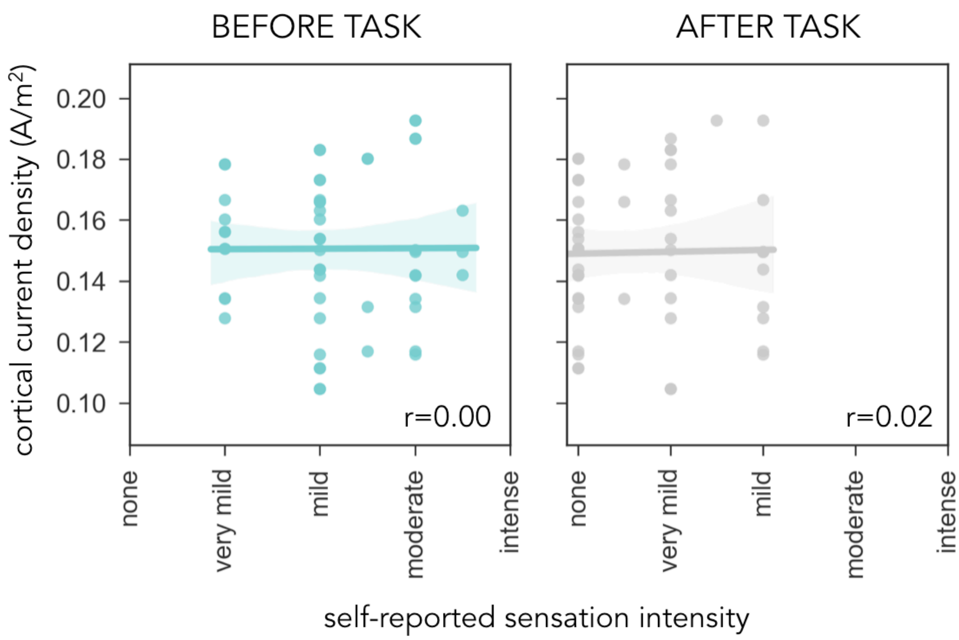

Figure 53: Relationship between self-reported stimulation intensities and estimated current densities at the cortical surface. Participants were asked to report the intensity of the stimulation 30 seconds after stimulation (before the task started). If participants finished the task during stimulation, they were asked to rate the intensity of the stimulation again. Only data from the active tDCS sessions are included in the plots.

A second interpretation is that the LH-bias and RH-bias stimulation protocols resulted in similar changes in behavior because human reasoning does not have a hemispheric lateralization component. Although patient studies provide strong evidence for different reasoning biases in each hemisphere, it is possible that these results do not extend to healthy individuals. That is, it is possible that the observed patterns of lateralization in reasoning are artifacts of abnormal brain functioning, and are not present in healthy brains. Although plausible, this explanation is inconsistent with our fMRI results, which generally consisted of asymmetric activations that were in line with our predictions.

Another possibility is that our general framework of hemispheric lateralization in reasoning is accurate, but we stimulated inappropriate neural targets. This possibility is more in line with our fMRI results. We expected that the uncertainty contrast and the advance > button press contrast would be associated with greater activity in the left vs. right interior frontal gyrus but we observed the opposite. This finding may help explain the similarity between LH-bias and RH-bias stimulation since we targeted the bilateral inferior frontal gyri in both tDCS experiments. If it is the case that the left IFG and right IFG play similar roles in reasoning, it is unclear why the effects of anodal and cathodal stimulation did not cancel out. That is, enhancing neural activity in one region and inhibiting activity in another region that does the same thing should result in a net neutral effect and should not depend on the intensity of stimulation. However, we consistently found that behavior was modulated by greater intensities of RH-bias stimulation at the cortical surface. If it is the case that the left and right IFG contribute to reasoning similarly and there is no hemispheric asymmetry, the effects of anodal and cathodal stimulation would have to be asymmetric to account for our results.

The fMRI results suggest that the left middle frontal gyrus may serve as a better neural target than the left inferior frontal gyrus. In the future, the tDCS experiments could be replicated with a different electrode montage that targets the left middle frontal gyrus and right inferior frontal gyrus. However, we should also note that it is likely that the stimulation in our tDCS studies extended outside of the inferior frontal gyri into the middle and superior frontal gyri since we did not optimize the electrode configuration for focality. In the individual brain maps, it is apparent that the highest current density occurs in the inferior frontal gyrus, but current is delivered to the entire frontal lobe to some extent.

A third interpretation is that the commonly held view that anodal stimulation increases cortical excitability and cathodal stimulation decreases excitability is inaccurate and our stimulation set up did not simultaneously increase activity in one hemisphere and decrease it in the other hemisphere as intended. As discussed previously, Jamil et al., 2017 did not find that cathodal stimulation inhibited motor cortex at intensities of 1.5mA and 2mA, which we applied in our experiments. Jacobson, Koslowsky, and Lavidor (2012) reviewed the inhibitory and excitatory effects of tDCS and found that studies generally report that anodal stimulation increases excitability in motor cortex and cathodal stimulation decreases excitability in motor cortex, but this excitation/inhibition relationship does not generalize to non-motor cortical regions; when brain areas outside the motor cortex are stimulated, the effect sizes of behavioral measures are much larger for anodal stimulation than for cathodal stimulation. Furthermore, cathodal stimulation has asymmetric effects on inducing synaptic plasticity (Kronberg et al., 2017) and neuronal firing (Rahman et al., 2013; Lafon et al., 2017) compared to anodal stimulation. As evidence that unilateral and bilateral stimulation protocols may produce differential effects, Leite and colleagues (2018) found that anodal right IFG stimulation had different effects on attentional switching costs than bilateral anodal right IFG stimulation and cathodal left IFG stimulation.

A combination of the second and third explanations best account for the results we observed. If 1) the neuromodulatory effects of anodal stimulation are stronger than cathodal stimulation and 2) we stimulated the wrong neural target in the left hemisphere, then we would expect to see greater effects in the RH-bias stimulation condition than the LH-bias stimulation condition and the effects could go in the same direction. These are two very strong assumptions that are only weakly backed up by evidence. Additional experiments are necessary to determine the individual contributions of the left IFG and right IFG to reasoning and to determine if a modified bilateral stimulation protocol produces opposite changes in behavior under LH-bias stimulation vs. RH-bias stimulation.

C. Future directions

In both tDCS experiments, we attempted to bias activity toward one hemisphere by applying bilateral stimulation to the inferior frontal gyri. Based on a body of literature showing that anodal stimulation increases cortical excitability and cathodal stimulation decrease cortical excitability (reviewed in Nitsche et al., 2008 and Stagg & Nitsche, 2011), we simultaneously applied anodal and cathodal stimulation to the inferior frontal gyri to induce hemispheric asymmetry in healthy individuals. Recent evidence suggests that cathodal and anodal stimulation may not produce opposite and equal effects, especially in non-motor brain areas, suggesting that our stimulation protocols may not have induced hemispheric asymmetry as intended. It would be informative to repeat both tDCS experiments with a modified stimulation protocol that stimulates each hemisphere in isolation in the future. A within-subjects design with three sessions consisting of anodal left frontal stimulation, anodal right frontal stimulation, and sham stimulation could shed light on the individual contributions of each frontal lobe on reasoning while obviating the need to interpret the asymmetric effects of cathodal and anodal stimulation. A modified version of the 4x1 ring system (Datta et al., 2009) could be adopted to deliver anodal or cathodal stimulation to one frontal lobe without influencing the contralateral hemisphere. One disadvantage of the 4x1 ring electrode configuration is that it delivers much less current to brain targets than the bilateral 2-electrode configuration that we used. Before adopting the unilateral stimulation protocol, it would be imperative to ensure that the montages can induce 1 V/m electric fields in target areas with modeling. Another option would be to use a bilateral stimulation but make the cortical densities under the cathode orders of magnitude weaker than those under the anode by increasing the size of the cathode or the distance between multiple cathodes.

We controlled for differences in current density in each participant's brain by modeling current flow in each participant and including estimates of cortical current densities in our regression analyses. In the future, differences in current delivery can be controlled more directly by applying stimulation at different intensities for different participants. The electric fields induced at target brain sites for a given stimulation intensity could be estimated for every participant, and then the current intensities needed to induce the same electric field strength across participants could be computed for each participant. Although this would not obviate individual differences in cortical excitability, endogenous neural activity, or network configurations, it could reduce the variability that arises from differences in current delivery. One caveat of this approach is that some participants will likely require stimulation intensities greater than 2mA, which may introduce safety concerns and exceed the limits of tDCS stimulators (for example, the stimulator we used can deliver a maximum of 2.5mA). As an additional level of control, the neuromodulatory effects of tDCS could be verified after stimulation with fMRI or EEG and directly accounted for in the behavioral analyses.

Finally, it would be interesting to run an additional fMRI experiment in the future with the same structure as the state guessing task, but using the ballot measures and criminal court cases used in the second tDCS experiment. This would serve to replicate our neuroimaging results and identify brain areas that are recruited when reasoning about scenarios with greater real-world consequences.

D. Hemispheric asymmetry in reasoning

Lateralized processing in the brain is both prevalent and beneficial: it maximizes the use of cortical space by reducing redundancy (Corballis, 1989; Gazzaniga, 2000; Vallortigara, 2006; Hopkins and Cantalupo, 2008), enhances brain efficiency by supporting dual processing (Rogers, 2000; Rogers et al., 2004; Vallortigara, 2006), reduces interhemispheric conflict by allowing the hemispheres to operate in separate problem spaces (Corballis, 2003; Vallortigara, 2006), and increases processing speed (Ringo et al., 1994). Hemispheric lateralization of inferential reasoning may be especially beneficial. Dual reasoning strategies—one driven to reduce uncertainty and the other driven to resolve inconsistency—create a flexible, efficient, and balanced reasoning system. Cognitive modules in the left hemisphere, with their propensities to create explanations, bridge gaps, and infer causation, may be preferentially recruited in situations that require creativity and liberal inference making. Conversely, cognitive modules in the right hemisphere, with their tendencies to detect conflict, monitor explanations in a global context, and inhibit inappropriate inferences, may play a greater role in situations that necessitate caution and conservative reasoning. Simply biasing these reasoning systems in accordance with situational demands ensures reasoning is adaptive, sensible, and efficient. In a healthy brain, the different inferential capabilities of the hemispheres enhance reasoning by maximizing both explanatory power and plausibility.

One consistent finding across all three experiments is that there is likely a hemispheric lateralization component to processing belief-evidence conflicts, but not all types of disconfirmatory evidence. Our results suggest that the right hemisphere does not continually evaluate or monitor all incoming evidence. Instead, it seems to only be recruited when strong evidence disconfirms strongly held beliefs. This is consistent with the results from our meta-analyses in which we found that simplistic evaluative labels ("rule finding" and "statement verification") were not right-lateralized, but the conflict label was completely right-lateralized. Is it also consistent with the differences in belief updating in response to conflict that we observed between tDCS experiments 1 and 2. In the first tDCS experiment, we found that LH-bias stimulation was associated with greater belief updates after conflict, but we found that RH-bias was stimulation was associated with greater updates after conflict in the second experiment. The tasks used in tDCS experiments 1 and 2 both involved receiving evidence that disconfirmed certain beliefs, but the belief-evidence conflicts in the second tDCS experiment were much stronger due to the types of stimuli used. Differences in belief-evidence conflict intensity can also explain why we found an interaction between RH-bias stimulation and conflict for the ballot measures, some of which were associated with very strong prior beliefs, but not the court cases, which were usually associated with neutral prior beliefs. Finally, in the fMRI experiment, we found right-lateralized networks for our conflict contrasts, but left-lateralized networks for the disconfirmatory evidence contrast. Together, these results suggest that the right hemisphere plays a pivotal role in processing salient conflicts between strong evidence and strong beliefs, but does not play a specialized role in simplistic evaluative processing.

In summary, we found that anodal tDCS to the right IFG and cathodal tDCS to the left IFG changed participants' reasoning compared to sham stimulation. Participants collected more evidence, adopted higher evidentiary thresholds for making a final inference, made less certain guesses than ideal, and were less likely to adopt more extreme beliefs in response to belief-evidence conflict. Surprisingly, we did not find equal and opposite results for anodal left IFG stimulation and cathodal right IFG stimulation. It is unclear whether the effects of the RH-bias stimulation were driven by cathodal stimulation to the left IFG, anodal stimulation to the right IFG, or a combination of both. All three views are consistent with our predictions concerning hemispheric asymmetry in reasoning. Future unilateral tDCS studies are necessary to disentangle the individual contributions of anodal and cathodal stimulation on each hemisphere. However, we did find at least some evidence that inducing hemispheric asymmetry with tDCS changes participants' reasoning biases, suggesting that there is some component of hemispheric asymmetry in reasoning.

References

1. Altemeyer, B. (2002). Dogmatic behavior among students: Testing a new measure of dogmatism. The Journal of Social Psychology, 142(6), 713-721.

2. Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. Trends in cognitive sciences, 18(4), 177-185.

3. Barbey, A., & Patterson, R. (2011). Architecture of explanatory inference in the human prefrontal cortex. Frontiers in psychology, 2, 162.

4. Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., ... & Brunoni, A. R. (2016). Safety of transcranial direct current stimulation: evidence based update 2016. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 9(5), 641-661.

5. Bisiach, E., Rusconi, M. L., & Vallar, G. (1991). Remission of somatoparaphrenic delusion through vestibular stimulation. Neuropsychologia, 29(10), 1029-1031.

6. Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. Psychonomic bulletin & review, 10(3), 730-737.

7. Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. Trends in cognitive sciences, 9(7), 322-328.

8. Braun, C. M. (2007). Evolution of hemispheric specialisation of antagonistic systems of management of the body's energy resources. Laterality, 12(5), 397-427.

9. Braun, C. M., & Suffren, S. (2011). A general neuropsychological model of delusion. Cognitive neuropsychiatry, 16(1), 1-39.

10. Borckardt, J. J., Bikson, M., Frohman, H., Reeves, S. T., Datta, A., Bansal, V., ... & George, M. S. (2012). A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. The Journal of Pain, 13(2), 112-120.

11. Bortoletto, M., Pellicciari, M. C., Rodella, C., & Miniussi, C. (2015). The interaction with task-induced activity is more important than polarization: a tDCS study. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8(2), 269-276.

12. Brownell, H. H., Potter, H. H., Bihrle, A. M., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. Brain and language, 27(2), 310-321.

13. Cabral, M. E., Baltar, A., Borba, R., Galvão, S., Santos, L., Fregni, F., & Monte-Silva, K. (2015). Transcranial direct current stimulation: before, during, or after motor training?. Neuroreport, 26(11), 618-622.

14. Caplan, R., & Dapretto, M. (2001). Making sense during conversation: an fMRI study. Neuroreport, 12(16), 3625-3632.

15. Caparelli-Daquer, E. M., Zimmermann, T. J., Mooshagian, E., Parra, L. C., Rice, J. K., Datta, A., ... & Wassermann, E. M. (2012, August). A pilot study on effects of 4× 1 high-definition tDCS on motor cortex excitability. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (pp. 735-738). IEEE.

16. Chakraborty, D., Truong, D. Q., Bikson, M., Kaphzan, H. (2017). Neuromodulation of Axon Terminals. Cerebral Cortex, 1-9.

17. Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M. T., & Munakata, Y. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. PloS one, 7(2), e31546.

18. Coltheart, M. (2010). The neuropsychology of delusions. Annals of the New York Academy of Sciences, 1191(1), 16-26.

19. Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. Annual review of psychology, 62, 271-298.

20. Coltheart, M., Menzies, P., & Sutton, J. (2010). Abductive inference and delusional belief. Cognitive neuropsychiatry, 15(1-3), 261-287.

21. Conway, C. R., Bollini, A. M., Graham, B. G., Keefe, R. S., Schiffman, S. S., & McEvoy, J. P. (2002). Sensory acuity and reasoning in delusional disorder. Comprehensive psychiatry, 43(3), 175-178.

22. Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. Journal of Neuroscience, 22(11), 4563-4567.

23. Corballis, M. C. (1989). Laterality and human evolution. Psychological review, 96(3), 492.

24. Corballis, P. M. (2003). Visuospatial processing and the right-hemisphere interpreter. Brain and cognition, 53(2), 171-176.

25. Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., ... & Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain, 130(9), 2387-2400.

26. Craig, A. D. (2005). Forebrain emotional asymmetry: a neuroanatomical basis?. Trends in cognitive sciences, 9(12), 566-571.

27. Cunillera, T., Fuentemilla, L., Brignani, D., Cucurell, D., & Miniussi, C. (2014). A simultaneous modulation of reactive and proactive inhibition processes by anodal tDCS on the right inferior frontal cortex. PloS one, 9(11), e113537.

28. D'Argembeau, A., Cassol, H., Phillips, C., Balteau, E., Salmon, E., & Van der Linden, M. (2013). Brains creating stories of selves: the neural basis of autobiographical reasoning. Social cognitive and affective neuroscience, 9(5), 646-652.

29. Danckert, J., Stöttinger, E., Quehl, N., & Anderson, B. (2011). Right hemisphere brain damage impairs strategy updating. Cerebral Cortex, 22(12), 2745-2760.

30. Darby, R., & Prasad, S. (2016). Lesion-related delusional misidentification syndromes: a comprehensive review of reported cases. The Journal of neuropsychiatry and clinical neurosciences, 28(3), 217-222.

31. Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D., & Bikson, M. (2009). Gyri-precise head model of transcranial direct current stimulation: improved spatial focality using a ring electrode versus conventional rectangular pad. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 2(4), 201-207.

32. Deglin, V. L., & Kinsbourne, M. (1996). Divergent thinking styles of the hemispheres: How syllogisms are solved during transitory hemisphere suppression. Brain and cognition, 31(3), 285-307.

33. Devinsky, O. (2009). Delusional misidentifications and duplications Right brain lesions, left brain delusions. Neurology, 72(1), 80-87.

34. Ditye, T., Jacobson, L., Walsh, V., & Lavidor, M. (2012). Modulating behavioral inhibition by tDCS combined with cognitive training. Experimental brain research, 219(3), 363-368.

35. Drummond, N. M., Cressman, E. K., & Carlsen, A. N. (2017). Offline continuous theta burst stimulation over right inferior frontal gyrus and pre-supplementary motor area impairs inhibition during a go/no-go task. Neuropsychologia, 99, 360-367.

36. Dudley, R. E. J., John, C. H., Young, A. W., & Over, D. E. (1997). Normal and abnormal reasoning in people with delusions. British Journal of Clinical Psychology, 36(2), 243-258.

37. Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. Neuroimage, 59(3), 2349-2361.

38. Eisenacher, S., & Zink, M. (2017). Holding on to false beliefs: The bias against disconfirmatory evidence over the course of psychosis. Journal of behavior therapy and experimental psychiatry, 56, 79-89.

39. Esmaeilpour, Z., Marangolo, P., Hampstead, B. M., Bestmann, S., Galletta, E., Knotkova, H., & Bikson, M. (2017). Incomplete evidence that increasing current intensity of tDCS boosts outcomes. Brain stimulation.

40. Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. Journal of Cognitive Neuroscience, 18(3), 320-334.

41. Ferstl, E. C., Guthke, T., & von Cramon, D. Y. (2002). Text comprehension after brain injury: Left prefrontal lesions affect inference processes. Neuropsychology, 16(3), 292.

42. Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind?. Neuroimage, 17(3), 1599-1612.

43. Friese, U., Rutschmann, R., Raabe, M., & Schmalhofer, F. (2008). Neural indicators of inference processes in text comprehension: an event-related functional magnetic resonance imaging study. Journal of Cognitive Neuroscience, 20(11), 2110-2124.

44. Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. Neuropsychologia, 43(8), 1204-1213.

45. Fujiyama, H., Hyde, J., Hinder, M. R., Kim, S. J., McCormack, G. H., Vickers, J. C., & Summers, J. J. (2014). Delayed plastic responses to anodal tDCS in older adults. Frontiers in aging neuroscience, 6, 115.

46. Gazzaniga, M. S. (1989). Organization of the human brain. Science, 245(4921), 947-952.

47. Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition?. Brain, 123(7), 1293-1326.

48. Gazzaniga, M. S. (2008). Human: The science behind what makes us unique.

49. Gazzaniga, M. S., & Miller, M. B. (2009). The left hemisphere does not miss the right hemisphere. The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology, 261-70.

50. Gazzaniga, M. S., & Smylie, C. S. (1984). Dissociation of language and cognition: A psychological profile of two disconnected right hemispheres. Brain, 107(1), 145-153.

51. Geuzaine, C., & Remacle, J. F. (2009). Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities. International journal for numerical methods in engineering, 79(11), 1309-1331.

52. Gill, J., Shah-Basak, P. P., & Hamilton, R. (2015). It's the thought that counts: examining the task-dependent effects of transcranial direct current stimulation on executive function. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8(2), 253-259.

53. Goel, V. (2007). Anatomy of deductive reasoning. Trends in cognitive sciences, 11(10), 435-441.

54. Goel, V., Buchel, C., Frith, C., & Dolan, R. J. (2000). Dissociation of mechanisms underlying syllogistic reasoning. Neuroimage, 12(5), 504-514.

55. Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. Cognition, 87(1), B11-B22.

56. Goel, V., Tierney, M., Sheesley, L., Bartolo, A., Vartanian, O., & Grafman, J. (2006). Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. Cerebral cortex, 17(10), 2245-2250.

57. Goel, V., Vartanian, O., Bartolo, A., Hakim, L., Ferraro, A. M., Isella, V., ... & Nichelli, P. (2013). Lesions to right prefrontal cortex impair real-world planning through prematurecommitments. Neuropsychologia, 51(4), 713-724.

58. Gurin, L., & Blum, S. (2017). Delusions and the right hemisphere: a review of the case for the right hemisphere as a mediator of reality-based belief. The Journal of neuropsychiatry and clinical neurosciences, 29(3), 225-235.

59. Hecht, D., Walsh, V., & Lavidor, M. (2010). Transcranial direct current stimulation facilitates decision making in a probabilistic guessing task. Journal of Neuroscience, 30(12), 4241-4245.

60. Heise, K. F., Niehoff, M., Feldheim, J. F., Liuzzi, G., Gerloff, C., & Hummel, F. C. (2014). Differential behavioral and physiological effects of anodal transcranial direct current stimulation in healthy adults of younger and older age. Frontiers in aging neuroscience, 6, 146.

61. Hogeveen, J., Grafman, J., Aboseria, M., David, A., Bikson, M., & Hauner, K. K. (2016). Effects of high-definition and conventional tDCS on response inhibition. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 9(5), 720-729.

62. Hone-Blanchet, A., Edden, R. A., & Fecteau, S. (2016). Online effects of transcranial direct current stimulation in real time on human prefrontal and striatal metabolites. Biological psychiatry, 80(6), 432-438.

63. Hopkins, W. D., & Cantalupo, C. (2008). Theoretical speculations on the evolutionary origins of hemispheric specialization. Current Directions in Psychological Science, 17(3), 233-237.

64. Horvath, J. C., Forte, J. D., & Carter, O. (2015). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8(3), 535-550.

65. Horvath, J. C., Forte, J. D., & Carter, O. (2015). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP amplitude modulation in healthy human subjects: a systematic review. Neuropsychologia, 66, 213-236.

66. Hsu, T. Y., Juan, C. H., & Tseng, P. (2016). Individual differences and state-dependent responses in transcranial direct current stimulation. Frontiers in human neuroscience, 10, 643.

67. Huang, Y., Liu, A. A., Lafon, B., Friedman, D., Dayan, M., Wang, X., ... & Parra, L. C. (2017). Measurements and models of electric fields in the in vivo human brain during transcranial electric stimulation. Elife, 6, e18834.

68. Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. The Quarterly Journal of Experimental Psychology, 40(4), 801-812.

69. Jacobson, L., Javitt, D. C., & Lavidor, M. (2011). Activation of inhibition: diminishing impulsive behavior by direct current stimulation over the inferior frontal gyrus. Journal of Cognitive Neuroscience, 23(11), 3380-3387.

70. Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. Experimental brain research, 216(1), 1-10.

71. Jamil, A., Batsikadze, G., Kuo, H. I., Labruna, L., Hasan, A., Paulus, W., & Nitsche, M. A. (2017). Systematic evaluation of the impact of stimulation intensity on neuroplastic after-effects induced by transcranial direct current stimulation. The Journal of physiology, 595(4), 1273-1288.

72. Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. Trends in cognitive sciences, 9(11), 512-518.

73. Kalbfleisch, M. L., Van Meter, J. W., & Zeffiro, T. A. (2007). The influences of task difficulty and response correctness on neural systems supporting fluid reasoning. Cognitive neurodynamics, 1(1), 71-84.

74. Kessler, S. K., Minhas, P., Woods, A. J., Rosen, A., Gorman, C., & Bikson, M. (2013). Dosage considerations for transcranial direct current stimulation in children: a computational modeling study. PloS one, 8(9), e76112.

75. Konishi, S., Hayashi, T., Uchida, I., Kikyo, H., Takahashi, E., & Miyashita, Y. (2002). Hemispheric asymmetry in human lateral prefrontal cortex during cognitive set shifting. Proceedings of the National Academy of Sciences, 99(11), 7803-7808.

76. Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. Annual review of psychology, 65.

77. Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. Brain research, 1243, 86-103.

78. Kronberg, G., Bridi, M., Abel, T., Bikson, M., & Parra, L. C. (2017). Direct current stimulation modulates LTP and LTD: activity dependence and dendritic effects. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 10(1), 51-58.

79. Kuo, H. I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Comparing cortical plasticity induced by conventional and high-definition 4× 1 ring tDCS: a neurophysiological study. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 6(4), 644-648.

80. Laakso, I., Tanaka, S., Koyama, S., De Santis, V., & Hirata, A. (2015). Inter-subject variability in electric fields of motor cortical tDCS. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 8(5), 906-913.

81. Lafon, B., Rahman, A., Bikson, M., & Parra, L. C. (2017). Direct current stimulation alters neuronal input/output function. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 10(1), 36-45.

82. Landmann, C., Dehaene, S., Pappata, S., Jobert, A., Bottlaender, M., Roumenov, D., & Le Bihan, D. (2006). Dynamics of prefrontal and cingulate activity during a reward-based logical deduction task. Cerebral Cortex, 17(4), 749-759.

83. Lauro, L. J. R., Rosanova, M., Mattavelli, G., Convento, S., Pisoni, A., Opitz, A., ... & Vallar, G. (2014). TDCS increases cortical excitability: Direct evidence from TMS–EEG. Cortex, 58, 99-111.

84. Leite, J., Gonçalves, Ó. F., Pereira, P., Khadka, N., Bikson, M., Fregni, F., & Carvalho, S. (2017). The differential effects of unihemispheric and bihemispheric tDCS over the inferior frontal gyrus on proactive control. Neuroscience research.

85. Li, L. M., Uehara, K., & Hanakawa, T. (2015). The contribution of interindividual factors to variability of response in transcranial direct current stimulation studies. Frontiers in cellular neuroscience, 9, 181.

86. Lie, C. H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. Neuroimage, 30(3), 1038-1049.

87. Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., & Hommer, D. W. (1999). Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. Journal of clinical and experimental neuropsychology, 21(1), 2-16.

88. Luft, C. D. B., Zioga, I., Banissy, M. J., & Bhattacharya, J. (2017). Relaxing learned constraints through cathodal tDCS on the left dorsolateral prefrontal cortex. Scientific reports, 7(1), 2916.

89. Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. Cognition, 124(1), 36-49.

90. Marcel, A. J., Tegnér, R., & Nimmo-Smith, I. (2004). Anosognosia for plegia: Specificity, extension, partiality and disunity of bodily unawareness. Cortex, 40(1), 19-40.

91. Marinsek, N., Turner, B. O., Gazzaniga, M., & Miller, M. B. (2014). Divergent hemispheric reasoning strategies: reducing uncertainty versus resolving inconsistency. Frontiers in human neuroscience, 8, 839.

92. Mayseless, N., & Shamay-Tsoory, S. G. (2015). Enhancing verbal creativity: modulating creativity by altering the balance between right and left inferior frontal gyrus with tDCS. Neuroscience, 291, 167-176.

93. McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. Research quarterly for exercise and sport, 60(1), 48-58.

94. McKenzie, C. R., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. Journal of Behavioral Decision Making, 15(1), 1-18.

95. McKay, R., Langdon, R., & Coltheart, M. (2007). Models of misbelief: Integrating motivational and deficit theories of delusions. Consciousness and Cognition, 16(4), 932-941.

96. Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2009). When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. Journal of Cognitive Neuroscience, 21(12), 2358-2368.

97. Miller, M. B., Kingstone, A., & Gazzaniga, M. S. (2002). Hemispheric encoding asymmetry is more apparent than real. Journal of Cognitive Neuroscience, 14(5), 702-708.

98. Miller, M. B., Sinnott-Armstrong, W., Young, L., King, D., Paggi, A., Fabri, M., ... & Gazzaniga, M. S. (2010). Abnormal moral reasoning in complete and partial callosotomy patients. Neuropsychologia, 48(7), 2215-2220.

99. Miller, M. B., & Valsangkar-Smyth, M. (2005). Probability matching in the right hemisphere. Brain and cognition, 57(2), 165-167.

100. Miranda, P. C., Lomarev, M., & Hallett, M. (2006). Modeling the current distribution during transcranial direct current stimulation. Clinical neurophysiology, 117(7), 1623-1629.

101. Monti, M. M., & Osherson, D. N. (2012). Logic, language and the brain. Brain research, 1428, 33-42.

102. Moritz, S., & Woodward, T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. British Journal of Clinical Psychology, 44(2), 193-207.

103. Mucci, A., Galderisi, S., Bucci, P., Tresca, E., Forte, A., Koenig, T., & Maj, M. (2005). Hemispheric lateralization patterns and psychotic experiences in healthy subjects. Psychiatry Research: Neuroimaging, 139(2), 141-154.

104. Muthalib, M., Besson, P., Rothwell, J., & Perrey, S. (2017). Focal Hemodynamic Responses in the Stimulated Hemisphere During High-Definition Transcranial Direct Current Stimulation. Neuromodulation: Technology at the Neural Interface.

105. Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., ... & Pascual-Leone, A. (2008). Transcranial direct current stimulation: state of the art 2008. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 1(3), 206-223.

106. Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. The Journal of physiology, 527(3), 633-639.

107. Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113.

108. Opitz, A., Paulus, W., Will, S., Antunes, A., & Thielscher, A. (2015). Determinants of the electric field during transcranial direct current stimulation. Neuroimage, 109, 140-150.

109. Parris, Ben A., Gustav Kuhn, Guy A. Mizon, Abdelmalek Benattayallah, and Tim L. Hodgson. "Imaging the impossible: an fMRI study of impossible causal relationships in magic tricks." Neuroimage 45, no. 3 (2009): 1033-1039.

110. Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). Schizophrenia bulletin, 30(4), 1005.

111. Phelps, E. A., & Gazzaniga, M. S. (1992). Hemispheric differences in mnemonic processing: The effects of left hemisphere interpretation. Neuropsychologia, 30(3), 293-297.

112. Rahman, A., Reato, D., Arlotti, M., Gasca, F., Datta, A., Parra, L. C., & Bikson, M. (2013). Cellular effects of acute direct current stimulation: somatic and synaptic terminal effects. The Journal of physiology, 591(10), 2563-2578.

113. Ramachandran, V. S. (1996). The evolutionary biology of self-deception, laughter, dreaming and depression: some clues from anosognosia. Medical Hypotheses, 47(5), 347-362.

114. Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. Journal of Politics, 64(4), 1021-1044.

115. Reverberi, C., Shallice, T., D'Agostini, S., Skrap, M., & Bonatti, L. L. (2009). Cortical bases of elementary deductive reasoning: Inference, memory, and metadeduction. Neuropsychologia, 47(4), 1107-1116.

116. Ringo, J. L., Doty, R. W., Demeter, S., & Simard, P. Y. (1994). Time is of the essence: a conjecture that hemispheric specialization arises from interhemispheric conduction delay. Cerebral Cortex, 4(4), 331-343.

117. Robertson, E. M., Theoret, H., & Pascual-Leone, A. (2003). Studies in cognition: the problems solved and created by transcranial magnetic stimulation. Journal of Cognitive Neuroscience, 15(7), 948-960.

118. Rodriguez-Moreno, D., & Hirsch, J. (2009). The dynamics of deductive reasoning: an fMRI investigation. Neuropsychologia, 47(4), 949-961.

119. Rogers, L. J. (2000). Evolution of hemispheric specialization: advantages and disadvantages. Brain and language, 73(2), 236-253.

120. Rogers, L. J., Zucca, P., & Vallortigara, G. (2004). Advantages of having a lateralized brain. Proceedings of the Royal Society of London B: Biological Sciences, 271(Suppl 6), S420-S422.

121. Rosen, D. S., Erickson, B., Kim, Y. E., Mirman, D., Hamilton, R. H., & Kounios, J. (2016). Anodal tDCS to right dorsolateral prefrontal cortex facilitates performance for novice jazz improvisers but hinders experts. Frontiers in human neuroscience, 10, 579.

122. Roser, M. E., Fugelsang, J. A., Dunbar, K. N., Corballis, P. M., & Gazzaniga, M. S. (2005). Dissociating processes supporting causal perception and causal inference in the brain. Neuropsychology, 19(5), 591.

123. Saucedo Marquez, C. M., Zhang, X., Swinnen, S. P., Meesen, R., & Wenderoth, N. (2013). Task-specific effect of transcranial direct current stimulation on motor learning. Frontiers in human neuroscience, 7, 333.

124. Schade, S., Moliadze, V., Paulus, W., & Antal, A. (2012). Modulating neuronal excitability in the motor cortex with tDCS shows moderate hemispheric asymmetry due to subjects' handedness: a pilot study. Restorative neurology and neuroscience, 30(3), 191-198.

125. Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. Proceedings of the National Academy of Sciences, 109(42), 17058-17062.

126. Simard, F., Joanette, Y., Petrides, M., Jubault, T., Madjar, C., & Monchi, O. (2010). Fronto-striatal contribution to lexical set-shifting. Cerebral Cortex, 21(5), 1084-1093.

127. Speechley, W. J., Ngan, E. T. C., Moritz, S., & Woodward, T. S. (2012). Impaired evidence integration and delusions in schizophrenia. Journal of Experimental Psychopathology, 3(4), 688-701.

128. Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. The Neuroscientist, 17(1), 37-53.

129. Stollstorff, M., Vartanian, O., & Goel, V. (2012). Levels of conflict in reasoning modulate right lateral prefrontal cortex. Brain research, 1428, 24-32.

130. Stone, S. P., Halligan, P. W., & Greenwood, R. J. (1993). The incidence of neglect phenomena and related disorders in patients with an acute right or left hemisphere stroke. Age and Ageing, 22(1), 46-52.

131. Stramaccia, D. F., Penolazzi, B., Sartori, G., Braga, M., Mondini, S., & Galfano, G. (2015). Assessing the effects of tDCS over a delayed response inhibition task by targeting the right inferior frontal gyrus and right dorsolateral prefrontal cortex. Experimental brain research, 233(8), 2283-2290.

132. Stagg, C. J., Lin, R. L., Mezue, M., Segerdahl, A., Kong, Y., Xie, J., & Tracey, I. (2013). Widespread modulation of cerebral perfusion induced during and after transcranial direct current stimulation applied to the left dorsolateral prefrontal cortex. Journal of Neuroscience, 33(28), 11425-11431.

133. Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome?. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481), 901-915.

134. Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., ... & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior

brain damage: effects of lesion location and test structure on separable cognitive processes. Neuropsychologia, 38(4), 388-402.

135. Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. Political Behavior, 31(2), 137-155.

136. Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. American Journal of Political Science, 50(3), 755-769.

137. Thielscher, A., Antunes, A., & Saturnino, G. B. (2015, August). Field modeling for transcranial magnetic stimulation: a useful tool to understand the physiological effects of TMS?. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE (pp. 222-225). IEEE.

138. Toner, K., Leary, M. R., Asher, M. W., & Jongman-Sereno, K. P. (2013). Feeling superior is a bipartisan issue: Extremity (not direction) of political views predicts perceived belief superiority. Psychological Science, 24(12), 2454-2462.

139. Tremblay, S., Lepage, J. F., Latulipe-Loiselle, A., Fregni, F., Pascual-Leone, A., & Théoret, H. (2014). The uncertain outcome of prefrontal tDCS. Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation, 7(6), 773-783.

140. Truong, D. Q., Magerowski, G., Blackburn, G. L., Bikson, M., & Alonso-Alonso, M. (2013). Computational modeling of transcranial direct current stimulation (tDCS) in obesity: impact of head fat and dose guidelines. NeuroImage: Clinical, 2, 759-766.

141. Tsujii, T., Masuda, S., Akiyama, T., & Watanabe, S. (2010). The role of inferior frontal cortex in belief-bias reasoning: an rTMS study. Neuropsychologia, 48(7), 2005-2008.

142. Tsujii, T., Sakatani, K., Masuda, S., Akiyama, T., & Watanabe, S. (2011). Evaluating the roles of the inferior frontal gyrus and superior parietal lobule in deductive reasoning: an rTMS study. Neuroimage, 58(2), 640-646.

143. Turner, B. O., Marinsek, N., Ryhal, E., & Miller, M. B. (2015). Hemispheric lateralization in reasoning. Annals of the New York Academy of Sciences, 1359(1), 47-64.

144. Vallortigara, G. (2006). The evolutionary psychology of left and right: costs and benefits of lateralization. Developmental psychobiology, 48(6), 418-427.

145. Vartanian, O., & Goel, V. (2005). Task constraints modulate activation in right ventral lateral prefrontal cortex. Neuroimage, 27(4), 927-933.

146. Villamar, M. F., Volz, M. S., Bikson, M., Datta, A., DaSilva, A. F., & Fregni, F. (2013). Technique and considerations in the use of 4x1 ring high-definition transcranial direct current stimulation (HD-tDCS). Journal of visualized experiments: JoVE, (77).

147. Vöröslakos, M., Takeuchi, Y., Brinyiczki, K., Zombori, T., Oliva, A., Fernández-Ruiz, A., ... & Berényi, A. (2018). Direct effects of transcranial electric stimulation on brain circuits in rats and humans. Nature communications, 9(1), 483.

148. Warman, D. M., Lysaker, P. H., Martin, J. M., Davis, L., & Haudenschield, S. L. (2007). Jumping to conclusions and the continuum of delusional beliefs. Behaviour research and therapy, 45(6), 1255-1269.

149. Wende, K. C., Nagels, A., Blos, J., Stratmann, M., Chatterjee, A., Kircher, T., & Straube, B. (2013). Differences and commonalities in the judgment of causality in physical and social contexts: an fMRI study. Neuropsychologia, 51(13), 2572-2580.

150. Wolford, G., Miller, M. B., & Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. Journal of Neuroscience, 20(6), RC64-RC64.

151. Woodward, T. S., Moritz, S., Cuttler, C., & Whitman, J. C. (2006). The contribution of a cognitive bias against disconfirmatory evidence (BADE) to delusions in schizophrenia. Journal of Clinical and Experimental Neuropsychology, 28(4), 605-617.

152. Xue, G., Juan, C. H., Chang, C. F., Lu, Z. L., & Dong, Q. (2012). Lateral prefrontal cortex contributes to maladaptive decisions. Proceedings of the National Academy of Sciences, 109(12), 4401-4406.

Appendix

Table A: Subjective reports of the effects of stimulation in the first tDCS experiment.

ID	LH-bias	SHAM	RH-bias
101	felt slightly dazed, high-like	no changes	harder to concentrate
102	felt more sure of decision	no changes	no changes
103	no changes, zoned out occasionally	no changes	no changes
104	hard to maintain focus, information seemed muddled	mentally hazy for a little bit	---
105	---	---	---
106	---	---	---
107	---	task seemed harder, harder to decide	---
108	harder to concentrate, harder to remember previous residents	---	---
109	no changes	no changes	hard to concentrate, harder to remember residents
110	---	lost focus a little, forgot decision during forced choice	felt like residents were passing by faster
111	---	no changes	no changes
112	less patient with the residents, would decide quickly, "good enough"	"caffeine buzz", slightly light-headed, cognition same	---
113	---	did it faster, especially at the end	---
114	slightly more distracted	felt tired	no changes
115	easily distracted	concentrating was more difficult	no changes
116	before would not move slider all the way, but this time moved it all the way toward ends, "go big or go home"	felt like could focus less	seemed easier, more clear
117	felt more unsure, especially for CA & NM	no changes	no changes
118	no changes	no changes	no changes
119	no changes	felt more impatient, sometimes forget	more decisive after seeing 4 people
120	no changes	felt like it went faster	no changes
121	felt a little more unable to concentrate	felt more concentrated	felt more closely related
122	no changes	felt like it was going faster	felt more tired
123	even more careful than last time	focus would vary, was more careful	felt like residents were going by quicker than last time
124	---	harder to focus	got really tired, harder to focus
125	no changes	no changes	felt easier, quicker. harder to remember previous residents
126	distracted in beginning because of tingling	getting sleepy	no changes